

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:23 ; Search time 2331.3 Seconds  
(without alignments)  
1578.034 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416

Perfect score: 223

Sequence: 1 ctaccataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*\*

1: gb\_ba:\*\*

2: gb\_htg:\*\*

3: gb\_in:\*\*

4: gb\_om:\*\*

5: gb\_ov:\*\*

6: gb\_pat:\*\*

7: gb\_ph:\*\*

8: gb\_pl:\*\*

9: gb\_pr:\*\*

10: gb\_ro:\*\*

11: gb\_sts:\*\*

12: gb\_sy:\*\*

13: gb\_un:\*\*

14: gb\_vi:\*\*

15: em\_ba:\*\*

16: em\_fun:\*\*

17: em\_hum:\*\*

18: em\_in:\*\*

19: em\_om:\*\*

20: em\_or:\*\*

21: em\_ov:\*\*

22: em\_pat:\*\*

23: em\_ph:\*\*

24: em\_pl:\*\*

25: em\_ro:\*\*

26: em\_sts:\*\*

27: em\_sy:\*\*

28: em\_un:\*\*

29: em\_vi:\*\*

30: em\_htgo\_hum:\*\*

31: em\_htgo\_inv:\*\*

32: em\_htgo\_rod:\*\*

33: em\_htg\_hum:\*\*

34: em\_htg\_inv:\*\*

35: em\_htg\_rod:\*\*

36: em\_htg\_other:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	223	100.0	416	6	AX127757	AX127757 Sequence
2	189.8	85.1	415	6	AX127755	AX127755 Sequence
3	47.8	21.4	186185	9	AC008817	AC008817 Homo sapi
4	46.4	20.8	110434	9	HSU66722	U66722 Human chrom
5	46	20.6	138350	2	AC010801	AC010801 Homo sapi
6	46	20.6	142126	2	AL596266	AL596266 Homo sapi
7	45.6	20.4	81503	2	AC091684	AC091684 Homo sapi
8	45.2	20.3	146349	2	AC007863	AC007863 Trypanoso
9	45.2	20.3	149241	2	AC068887	AC068887 Homo sapi
10	44.2	19.8	158236	9	AC011631	AC011631 Homo sapi
11	44.2	19.8	216431	2	AL354771	AL354771 Homo sapi
12	44	19.7	40917	6	AX033911	AX033911 Sequence
13	44	19.7	41008	6	AX033912	AX033912 Sequence
14	44	19.7	80155	9	AF042484	AF042484 Homo sapi
15	43.8	19.6	80622	9	AF222855	AF222855 Homo sapi
16	43.8	19.6	175959	2	AL450429	AL450429 Homo sapi
17	43.6	19.6	228	8	AY020821	AY020821 Oryza sat
18	43.6	19.6	234	8	AY018525	AY018525 Oryza sat
19	43.6	19.6	185499	2	AC090578	AC090578 Homo sapi
20	43.6	19.6	187967	2	AC026089	AC026089 Homo sapi
21	43.4	19.5	145358	2	AC091086	AC091086 Oryza sat
22	43.2	19.4	108523	9	AC000159	AC000159 Homo sapi
23	43.2	19.4	135237	2	AP000814	AP000814 Homo sapi
24	43.2	19.4	156736	9	AC006502	AC006502 Homo sapi
25	43	19.3	191962	2	AC048342	AC048342 Homo sapi
26	43	19.3	193996	2	AC068503	AC068503 Homo sapi
27	43	19.3	203718	9	AC009486	AC009486 Homo sapi
28	42.8	19.2	156060	2	AC004153	AC004153 Plasmodi
29	42.8	19.2	165114	2	AC011854	AC011854 Homo sapi
30	42.8	19.2	175020	9	AC022317	AC022317 Homo sapi
31	42.8	19.2	185702	2	AP002963	AP002963 Homo sapi
32	42.8	19.2	211316	9	AC009294	AC009294 Homo sapi
33	42.8	19.2	269661	2	AC023911	AC023911 Homo sapi
34	42.4	19.0	69058	9	AF222856	AF222856 Homo sapi
35	42.4	19.0	80202	9	AF222854	AF222854 Homo sapi
36	42.4	19.0	101584	9	CNS01DS5	AL121655 BAC seque
37	42.4	19.0	155390	2	AL596118	AL596118 Homo sapi
38	42.4	19.0	162958	2	AC011022	AC011022 Homo sapi
39	42.4	19.0	168519	2	AL355340	AL355340 Homo sapi
40	42.4	19.0	168671	2	AC019247	AC019247 Homo sapi
41	42.4	19.0	173435	2	AC015715	AC015715 Homo sapi
42	42.4	19.0	176432	2	AL391648	AL391648 Homo sapi
43	42.4	19.0	180942	33	AL512845	AL512845 Human DNA
44	42.2	18.9	66623	2	AC073039	AC073039 Homo sapi
45	42.2	18.9	110880	9	AC004878	AC004878 Homo sapi

ALIGNMENTS

RESULT 1  
AX127757  
LOCUS AX127757 416 bp DNA  
DEFINITION Sequence 10 from Patent WO0131042.  
ACCESSION AX127757 PAT 15-MAY-2001  
VERSION AX127757.1 GI:14134404  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
synthetic construct.  
artificial sequence.  
REFERENCE  
AUTHORS 1 (bases 1 to 416)  
TITLE Weston,B. and de Beuckeleer,M.  
JOURNAL Male-sterile brassica plants and methods for producing same  
Patent: WO 0131042-A 10 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source  
location/Qualifiers  
1..416  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="3' border flanking region of elite event MS-B2"

```

misc_feature      1..193
                  /note="T-DNA"
misc_feature      194..416
                  /note="plant DNA"
BASE COUNT      137 a      72 c      54 g      152 t      1 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 223; DB 6; Length 416;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ctaccataattataattataattataataactgaacacatggtgcccccctgctgcttta 60
Db      194 CTACCATTAATTATAATTATAATTATAATATACTGAACCATGGTGCCTGCTGCTTTA 253

QY      61 catggattctccgcctactatttgatcacggtgatatatataccgataataatgacatatatt 120
Db      254 CATGGATTCTCCGCTACTATTGTAACGGTGATATATACCGTATAATGTACATATATT 313

QY      121 tatatgaacatgattcaatgctgtgagttgttctcatccgtaagagtttcaatgatgaat 180
Db      314 TATATGAACATGATTAATGCTGTGTGAGTTGTTCTCATCCGTACAGGTTTCAATATGTAAT 373

QY      181 ggtgaagagtcacaaacccaaataatcatgaacacccaaactcgat 223
Db      374 GGTGAAGAGTCAAAACCCAAAATCATGAACACCCAAACTCGAT 416

RESULT 2
AX127755/c
LOCUS      AX127755      415 bp      DNA      PAT      15-MAY-2001
DEFINITION      Sequence 8 from Patent WO0131042.
ACCESSION      AX127755
VERSION      AX127755.1 GI:14134402
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
  1 (bases 1 to 415)
  Weston,B. and de Beuckeleer,M.
  Male-sterile brassica plants and methods for producing same
  Patent: WO 0131042-A 8 03-MAY-2001,
  Aventis CropScience N.V. (BE)
FEATURES
  source
    1..415
      /organism="synthetic construct"
      /db_xref="taxon:32630"
      /notes="5', border flanking region of elite event MS-B2"
misc_feature
    1..234
      /note="plant DNA"
misc_feature
    235..415
      /note="T-DNA"
BASE COUNT      154 a      55 c      70 g      136 t
ORIGIN

Query Match
Best Local Similarity 85.1%; Score 189.8; DB 6; Length 415;
Matches 213; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY      1 ctaccataattataattataattataataatactgaacacatggtgcccccctgctgcttt 59
Db      222 CTACCATTAATTATAATTATAATTATAATATACTTGAACCATGGTGCCTGCTGCTTT 163

QY      60 acatggattctccgcctactatttgatcacggtgatatatataccgataataatgacatatat 119
Db      162 ACATGGATTCTTCGCTACTATTGTAACGGTGATATATA-CGTATAATGTACATATAT 104

QY      120 ttatatgaacatgattcaatgctgtgagttgttctcatccgtaagagtttcaatgatgaat 179
Db      103 TTATATGAACATGATTAATGCTGTGTGAGTTGTTCTCATCCGTACAGGTTTCAATATGTAA 44

```



[illegible]



```
* 31449 31548: gap of 100 bp
* 31549 37912: contig of 6364 bp in length
* 37913 38012: gap of 100 bp
* 38013 44269: contig of 6257 bp in length
* 44270 44369: gap of 100 bp
* 44370 51461: contig of 7092 bp in length
* 51462 51561: gap of 100 bp
* 51562 59216: contig of 7655 bp in length
* 59217 59316: gap of 100 bp
* 59317 69884: contig of 10568 bp in length
* 69885 69984: gap of 100 bp
* 69985 80594: contig of 10610 bp in length
* 80595 80694: gap of 100 bp
* 80695 93129: contig of 12435 bp in length
* 93130 93228: gap of 100 bp
* 93230 106980: contig of 13751 bp in length
* 106981 107080: gap of 100 bp
* 107081 121789: contig of 14709 bp in length
* 121790 121889: gap of 100 bp
* 121890 138350: contig of 16461 bp in length.

FEATURES
  source
    1. .138350
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="RP11-2L13"
      /clone_lib="RPCI-11 Human Male BAC"
    1. .1169
      /note="assembly_fragment"
      1270. .4174
      /note="assembly_fragment"
      4275. .6879
      /note="assembly_fragment"
      6980. .10304
      /note="assembly_fragment"
      10405. .13133
      /note="assembly_fragment"
      13234. .16716
      /note="assembly_fragment"
      16817. .21929
      /note="assembly_fragment"
      22030. .26268
      /note="assembly_fragment"
      26369. .31448
      /note="assembly_fragment"
      clone_end:T7
      vector_side:right
      31549. .37912
      /note="assembly_fragment"
      38013. .44269
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:right
      44370. .51461
      /note="assembly_fragment"
      51562. .59216
      /note="assembly_fragment"
      59317. .69884
      /note="assembly_fragment"
      69985. .80594
      /note="assembly_fragment"
      80695. .93129
      /note="assembly_fragment"
      93230. .106980
      /note="assembly_fragment"
      107081. .121789
      /note="assembly_fragment"
      121890. .138350
      /note="assembly_fragment"
      42610 a 25076 c 24703 g 44151 t 1810 others

BASE COUNT
ORIGIN
```

Query Match

20.6%; Score 46; DB 2; Length 138350;

```
Best Local Similarity 54.0%; Pred No. 0.16;
Matches 94; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 10 ttataattataataataactgaaccatggtgccccctgctgttacatgattt 69
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130960 TTATGCATACACTTAGATATATATACATACAGATACAGATTTATATATATATATAT 131019

QY 70 ctccgctactattgtatatactggtatataaccgtataatgtacatatatttatgaac 129
| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 131020 ATATATATATATATATATATATATATATATATATATATATATATATAGGTATCC 131079

QY 130 atgattaatgcttgagtggttcttcacatcgtaagagtttcaatgttaatggt 183
| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131080 CTTCCTTATATAGGTACATGTACACACCTATATCTATATATATATATCTACAGTT 131133

RESULT 6
AL596266
LOCUS AL596266 142126 bp DNA HTG 29-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-2L13, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION AL596266
VERSION AL596266.2 GI:15028804
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 142126)
AUTHORS Melay, K.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
COMMENT On Jul 29, 2001 this sequence version replaced gi:15022382.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA2L13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 141907 bases at least Q40
Consensus quality: 141961 bases at least Q30
Consensus quality: 142003 bases at least Q20
Insert size: 142026; sum-of-contigs
Quality coverage: 12.33x in Q20 bases; sum-of-contigs Quality
coverage: 10.95x in Q20 bases; agarose-gp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6452: contig of 6452 bp in length
* 6453 6552: gap of 100 bp
* 6553 142126: contig of 135574 bp in length.

FEATURES
  source
    1. .142126
      Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /clone="RP11-2L13"
        /clone_lib="RPCI-11.1"
    1. .6452
      misc_feature
```



★ 78989 81503: contig of 2515 bp in length.

[illegible]

```
BASE COUNT   37892 a 34490 c 35232 g 38734 t      1 others
ORIGIN

Query Match      20.3%;   Score 45.2;   DB 2;   Length 146349;
Best Local Similarity 54.9%;   Pred. No. 0.25;
Matches 89;   Conservative 0;   Mismatches 73;   Indels 0;   Gaps 0;

QY  2  taccataattataattataattataactgaacaccactggtgccccctgctgtttac 61
    || || || || || || || || || || || || || || || || || || || || ||
Db  11970 TATTATTATTATAATTAACGTGTTGGTATTCATCGACTTGCTACCCCTGGGTATT 11911

QY  62 atggattctccgctactattgtatatacgtgtatataacgtatataatgtacatatatt 121
    || || || || || || || || || || || || || || || || || || || || ||
Db  11910 CTATATTTCTATATTTCTATATGCATATATATATATATATATATTTGTGTGTCGTGCT 11851

QY  122 atatgaacatgatgaatgctgtgctgtggtgttctcctacgttaa 163
    || || || || || || || || || || || || || || || || || || || || ||
Db  11850 GTGTGTTGACGTTGCTGCTGTTGTTGCTGCTGCTGCTGTTGTTGTTA 11809

RESULT  9
AC068887
LOCUS      AC068887      149241 bp      DNA      HWC      09-AUG-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-392G11, WORKING DRAFT
ACCESSION  AC068887
VERSION    AC068887.2  GI:15145500
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 149241)
            Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amarakunte,H.C., Are,J.R., Banks,T., Barbra,J.,
            Benton,J., Blimie,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louisgied,H., Lozardo,R.J., Lu,X., Lucier,A., Lucier,B., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Picken,R., Pridus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Roife,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
            Sisson,I., Sodergren,E., Sonaik,I., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
            Thomas,S., Umanik,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Washington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstein,G. and Gibbs,R.
```

```
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 149241)
AUTHORS    Worley,K.C.
TITLE      Direct Submission
JOURNAL    Submitted (11-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Aug 9, 2001 this sequence version replaced gi:14717197.
COMMENT    -----
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -----
            Project Information
            Center project name: HAOU
            Center clone name: RP11-392G11
            -----
            Summary Statistics
            Sequencing vector: Plasmid; M7789
            Sequencing vector: M13; L08821
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 140201 bases at least Q40
            Consensus quality: 145535 bases at least Q30
            Consensus quality: 147437 bases at least Q20
            Estimated insert size: 158898; sum-of-ctngs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 5.9x in Q20 bases; sum-of-ctngs estimation
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 38028: contig of 38028 bp in length
            * 38029 38128: gap of unknown length
            * 38129 61953: contig of 23825 bp in length
            * 61954 62053: gap of unknown length
            * 62054 85448: contig of 23395 bp in length
            * 85449 85548: gap of unknown length
            * 85549 105941: contig of 20393 bp in length
            * 105942 122966: contig of 16925 bp in length
            * 122967 123066: gap of unknown length
            * 123067 133765: contig of 10699 bp in length
            * 133766 133865: gap of unknown length
            * 133866 142256: contig of 8391 bp in length
            * 142257 142356: gap of unknown length
            * 142357 146773: contig of 4417 bp in length
            * 146774 146874: gap of unknown length
            * 146874 149241: contig of 2368 bp in length.
            *
            * Location/Qualifiers
            * 1. 149241
            * /organism="Homo sapiens"
            * /db_xref="taxon:9606"
            * /chromosome="12"
            * /clone="RP11-392G11"
            *
            BASE COUNT  42918 a 32275 c 31743 g 41487 t      818 others
            ORIGIN

Query Match      20.3%;   Score 45.2;   DB 2;   Length 149241;
Best Local Similarity 54.1%;   Pred. No. 0.25;
Matches 92;   Conservative 0;   Mismatches 78;   Indels 0;   Gaps 0;

QY  6  ataattataattataattataactgaacaccactggtgccccctgctgtttacatgg 65
    || || || || || || || || || || || || || || || || || || || || ||
Db  106625 ATATATATATGTTATATATGCGCATATATACATATATATACATATATATACATAT 106684
```







Query Match		19.7% ; Score 44 ; DB 6 ; Length 40917 ;	
Best Local Similarity		58.3% ; Pred. No. 0.52 ;	
Matches		77 ; Conservative 0 ; Mismatches 55 ; Indels 0 ; Gaps 0 ;	
QY	5	cataattataattataataactgaacaccatggtgccccctgctgctttacatg	64
Db	15579	CATAAATATATACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTG	15638
QY	65	gattctccgctactattgtatatacgtgtatataccgtataatgtacatatatttata	124
Db	15639	TATATATACATAAATATGTATATGTGTATATAGACATAAATATGTATATGTGTATA	15698
QY	125	tgaacatgatta	136
Db	15699	TAGACATAAATA	15710
RESULT	13		
AX033912			
LOCUS	AX033912	41008 bp	DNA
DEFINITION	Sequence 4 from Patent WO9851790.		PAT
ACCESSION	AX033912		
VERSION	AX033912.1	GI:10280480	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unidentified		
	unclassified.		
REFERENCE	1 (bases 1 to 41008)		
AUTHORS	Cancilla,M.R., Choo,K.H. and Du,S.D.		
TITLE	A novel nucleic acid molecule		
JOURNAL	Patent: WO 9851790-A 4 19-NOV-1998;		
	CANCILLA MICHAEL ROBERT (AU) ; CHOO KONG HONG ANDY (AU) ; SART		
	DESIREE DU (AU) ; AMRAD OPERATIONS PTY LTD (AU)		
FEATURES	Location/Qualifiers		
	source		
	1..41008		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	11778 a	8413 c	8415 g
ORIGIN	12398 t	4	others
Query Match		19.7% ; Score 44 ; DB 6 ; Length 41008 ;	
Best Local Similarity		58.3% ; Pred. No. 0.52 ;	
Matches		77 ; Conservative 0 ; Mismatches 55 ; Indels 0 ; Gaps 0 ;	
QY	5	cataattataattataataactgaacaccatggtgccccctgctgctttacatg	64
Db	15314	CATAAATATATACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTG	15373
QY	65	gattctccgctactattgtatatacgtgtatataccgtataatgtacatatatttata	124
Db	15374	TATATATACATAAATATGTATATATGTGTATATAGACATAAATATGTATATGTGTATA	15433
QY	125	tgaacatgatta	136
Db	15434	TAGACATAAATA	15445
RESULT	14		
AX042484			
LOCUS	AX042484	80155 bp	DNA
DEFINITION	Homo sapiens sequence of the 80 kb core centromere protein binding		PRI
	region of the mar del(10) neocentromere corresponding to human		
	chromosome 10q25.2, complete sequence.		
ACCESSION	AF042484		
VERSION	AF042484.1	GI:4205782	
KEYWORDS			
SOURCE	HTG.		
	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 80155)		
AUTHORS	Barry,A.E., Howman,E.V., Cancilla,M.R., Saffery,R. and Choo,K.H.A.		
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/chromosome="1"			
/map="q25.11-31.1"			
/clone="RP11-545P21"			
/clone_lib="RPC1-11.2"			
1..18609			
/note="assembly_fragment:02814			
clone_end:SP6			
vector_side:left"			
18710..27338			
/note="assembly_fragment:01738			
fragment_chain:1"			
27439..126572			
/note="assembly_fragment:01150			
fragment_chain:1"			
126673..206680			
/note="assembly_fragment:02306			
fragment_chain:1"			
206781..216431			
/note="assembly_fragment:00386			
fragment_chain:1			
clone_end:T7			
vector_side:right"			
BASE COUNT	70588 a	35565 c	36468 g
ORIGIN	73389 t	421	others
Query Match		19.8% ; Score 44.2 ; DB 2 ; Length 216431 ;	
Best Local Similarity		50.2% ; Pred. No. 0.42 ;	
Matches		109 ; Conservative 0 ; Mismatches 108 ; Indels 0 ; Gaps 0 ;	
QY	6	ataattataattataataactgaacaccatggtgccccctgctgctttacatg	65
Db	54915	ATGTTCTATGGGAACGGTAAATAATATGACCCAGTGGCATGATAAGATATATATAT	54974
QY	66	attctccgctactattgtatacgtgtatataccgtataatgtacatatatttata	125
Db	54975	ATATGTATGTGTAT	55034
QY	126	gaacatgattaact	

TITLE Sequence analysis of an 80 kb human neocentromere  
JOURNAL Hum. Mol. Genet. 8 (2), 217-227 (1999)  
MEDLINE 99135901  
PUBMED 9931329  
REFERENCE 2 (bases 1 to 80155)  
AUTHORS Barry,A.E., Howman,E.V., Cancelli,M.R., Saffery,R. and Choo,A.  
TITLE Direct Submission  
JOURNAL Submitted (13-JAN-1998) Chromosome Research Unit, The Murdoch  
Institute, 10th Floor, Royal Childrens Hospital, Flemington Rd.,  
Parkville, Melbourne, Victoria 3052, Australia  
FEATURES  
source Location/Qualifiers  
1..80155  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/map="10q25.2"  
/note="Sequence information obtained from BAC clones  
prepared from cell line BE2Cl-18-5f"  
complement(join(49..525,3628..4545,4849..5053))  
repeat\_region  
/rpt\_family="LIMEC"  
/rpt\_type=dispersed  
join(560..597,3184..3281)  
repeat\_region  
/rpt\_family="MER58"  
/rpt\_type=dispersed  
complement(join(598..2553,2716..3183))  
repeat\_region  
/rpt\_family="TIGER1"  
/rpt\_type=dispersed  
2558..2581  
/standard\_name="T1"  
/rpt\_type=direct  
/rpt\_unit=(CAGGCACAGTGG)2  
repeat\_region  
2570..2715  
/rpt\_family="AluSg/x"  
/rpt\_type=dispersed  
3021..3050  
/standard\_name="T2"  
/note="83.3% homology"  
/rpt\_type=direct  
/rpt\_unit=(TACACAAAGTG)3  
repeat\_region  
3283..3558  
/rpt\_family="AluSx"  
/rpt\_type=dispersed  
4173..4211  
/standard\_name="T3"  
/note="82.1% homology"  
/rpt\_type=direct  
/rpt\_unit=(AAAAAATAATT)3  
repeat\_region  
complement(4546..4848)  
/rpt\_family="AluSx"  
/rpt\_type=dispersed  
complement(5909..6208)  
repeat\_region  
/rpt\_family="AluSg"  
/rpt\_type=dispersed  
join(6232..6573,6638..6688)  
repeat\_region  
/rpt\_family="MLT2CB"  
/rpt\_type=dispersed  
6598..6645  
/standard\_name="AC24"  
/note="100% homology"  
/rpt\_type=tandem  
/rpt\_unit=AC  
6598..6637  
/note="microsatellite"  
repeat\_region  
/rpt\_type=tandem  
/rpt\_unit=CA  
7091..7134  
/rpt\_family="MIR"  
/rpt\_type=dispersed  
7808..7911  
/rpt\_family="L2"  
/rpt\_type=dispersed  
8349..8524

/rpt\_family="MIR"  
/rpt\_type=dispersed  
complement(8839..9124)  
/rpt\_family="AluJb"  
/rpt\_type=dispersed  
9180..9280  
/rpt\_family="MIR"  
/rpt\_type=dispersed  
10070..10170  
/note="low complexity"  
/rpt\_family="purine-rich"  
10097..10136  
/standard\_name="GA1"  
/note="82.5% homology"  
/rpt\_type=direct  
/rpt\_unit=(GAAAGAAAGG)4  
10175..10339  
/rpt\_family="FRAM"  
/rpt\_type=dispersed  
complement(10899..11197)  
/rpt\_family="AluSx"  
/rpt\_type=dispersed  
join(11627..11757,12434..12512,13443..13562)  
/note="98% identity"  
/db\_xref="dbEST:AA774571"  
11996..12028  
/standard\_name="AT4"  
/note="100% homology"  
/rpt\_type=tandem  
/rpt\_unit=(ATTAT)8  
complement(12010..12312)  
/rpt\_family="AluJb"  
/rpt\_type=dispersed  
12327..12433  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(13946..14235)  
/rpt\_family="AluSg"  
/rpt\_type=dispersed  
15130..15159  
/note="low complexity"  
/rpt\_family="AT-rich"  
15164..15753  
/standard\_name="AT28"  
/note="82.1% homology"  
/rpt\_type=direct  
/rpt\_unit=(ATGTATATATCTGTATATAGACATAAAAT)21  
join(15170..15289,15410..15612)  
/note="microsatellite"  
/rpt\_type=tandem  
/rpt\_unit=TA  
15617..15731  
/note="microsatellite"  
/rpt\_type=tandem  
/rpt\_unit=CATA  
16004..16100  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(16128..16364)  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(16407..16592)  
/rpt\_family="L2"  
/rpt\_type=dispersed  
complement(16657..16889)  
/rpt\_family="MIR"  
/rpt\_type=dispersed  
16914..16936  
/note="low complexity"  
/rpt\_family="AT-rich"  
17102..17386  
/rpt\_family="AluSx"  
/rpt\_type=dispersed





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:36 ; Search time 716.55 Seconds  
(without alignments)  
266.811 Million cell updates/sec

Title: us-09-698-903b-10\_COPY\_194\_416

Perfect score: 223

Sequence: 1 ctacataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930631 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	100.0	416	22	Left (3') border f
2	189.8	85.1	415	22	Right (5') border
3	50.4	22.6	244	22	Oligonucleotide D1
4	50	22.4	936	22	Oligonucleotide D1
5	50	22.4	936	22	Oligonucleotide D1
6	50	22.4	936	22	Oligonucleotide D1
7	50	22.4	936	22	Oligonucleotide D1
8	50	22.4	936	22	Oligonucleotide D1
9	50	22.4	936	22	Oligonucleotide D1
10	50	22.4	936	22	Oligonucleotide D1
11	50	22.4	936	22	Oligonucleotide D2

12	50	22.4	936	22	AAF58262	Oligonucleotide D2
13	50	22.4	936	22	AAF58262	Oligonucleotide D2
14	50	22.4	938	22	AAF58255	Oligonucleotide D1
15	50	22.4	938	22	AAF58255	Oligonucleotide D1
16	47.4	21.3	244	22	AAF58238	Oligonucleotide D1
17	44	19.7	1153	21	AAC39525	Arabidopsis thalia
18	44	19.7	80240	20	AAV83940	NC-contig derived
19	44	19.7	80595	20	AAV83939	NC-contig derived
20	40.8	18.3	1152	21	AAC46330	Arabidopsis thalia
21	38.8	17.4	513445	22	AAI61373	Soybean 318013 reg
22	38.2	17.1	1119	8	AAV71222	Sequence of soybean
23	37.4	16.8	32190	22	AAI62927	Human genomic DNA
24	37	16.6	470	22	AAH93301	Plasmodium falcipa
25	37	16.6	50000	21	AAA64139	Nucleotide sequenc
26	36.8	16.5	426	22	AAH93297	Arabidopsis thalia
27	36.8	16.5	15577	19	AAV35616	SHOX gene prelinin
28	36.8	16.5	32367	19	AAV35620	SHOX gene prelinin
29	36.6	16.4	893	22	AAH03844	Human SHOX (short
30	36.6	16.4	1602	22	AAH15369	Human CDNA clone (
31	36.4	16.3	2418	13	AAQ27886	Human CDNA sequenc
32	36.4	16.3	32249	22	AAI62932	P.falciparum GBP13
33	36.2	16.2	412	22	AAH93300	Human genomic DNA
34	35.8	16.1	916	22	AAH05774	Plasmodium falcipa
35	35.8	16.1	1558	21	AAH1567	Arabidopsis thalia
36	35.8	16.1	335913	22	AAI61371	N. meningitidis pa
37	35.8	16.1	335913	22	AAI61372	Soybean 240017 reg
38	35	15.7	41100	22	AAH10873	Soybean 240017 reg
39	35	15.7	101786	21	AAV22293	Human genomic DNA
40	34.8	15.6	15418	21	AAH3785	BAC containing rep
41	34.8	15.6	16875	22	AAH21613	Nucleotide sequenc
42	34.8	15.6	16875	22	AAH21613	Human hypocrerin r
43	34.8	15.5	6038	21	AAZ43720	Human FSH-beta DNA
44	34.6	15.5	7622	21	AAZ43718	Human FSH-beta DNA
45	34.4	15.4	2451	20	AAV82439	Vascular endotheli

ALIGNMENTS

RESULT 1

AAD06999

ID AAD06999 standard: DNA; 416 BP.

XX

AC AAD06999;

XX

DT 06-AUG-2001 (first entry)

XX

DE Left (3') border flanking region of elite event MS-B2.

XX

KW MS-B2 elite event; transgenic Brassica plant; transformation event;

KW male-sterility gene; ds.

XX

OS Chimeric - Agrobacterium sp.

OS Chimeric - Brassica sp.

XX

FH Key Location/Qualifiers

FT misc\_feature 1..193

FT /tag= a

FT /note= "Corresponds to T-DNA"

FT misc\_feature 194..416

FT /tag= b

FT /note= "Corresponds to plant DNA"

XX

XX WO200131042-A2.

XX

XX 03-MAY-2001.

 XX | 26-OCT-2000; 2000WO-EP10680. |



[illegible]















**THIS PAGE BLANK (USP 10)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:29 ; Search time 301.6 Seconds  
(without alignments)  
167.456 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416

Perfect score: 223

Sequence: 1 ctaccataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	19.7	80246	4	US-09-078-294-4
2	44	19.7	80595	4	US-09-078-294-3
c 3	34.6	15.5	6038	4	US-09-305-639-4
c 4	34.6	15.5	7622	4	US-09-305-639-1
5	33.2	14.9	1160	1	US-08-006-082A-4
6	33.2	14.9	1264	1	US-08-006-082A-1
c 7	32.8	14.7	1951	4	US-09-058-947A-12
c 8	32.8	14.7	1983	1	US-08-383-756-3
c 9	32.8	14.7	1983	2	US-08-460-898-3
c 10	32.8	14.7	3722	4	US-09-058-947A-1
11	32.4	14.5	51952	3	US-08-947-823-1
c 12	32	14.3	4254	2	US-08-443-639-7
c 13	32	14.3	8920	2	US-08-446-855A-1
c 14	32	14.3	8920	4	US-09-150-741-1
15	31.4	14.1	480	1	US-08-282-581-4
16	31.4	14.1	480	1	US-08-550-544-4
c 17	31.4	14.1	1601	3	US-08-817-188-3
18	31.4	14.1	2799	2	US-08-874-186-42
19	31.4	14.1	3627	1	US-08-104-072B-6
20	31.4	14.1	3627	1	US-08-351-413-7
21	31.4	14.1	3627	2	US-09-025-583-7
c 22	31.2	14.0	19124	2	US-08-487-826B-13
c 23	30.8	13.8	5993	4	US-09-383-630-1
c 24	30.8	13.8	5993	4	US-09-383-630-2
c 25	30.6	13.7	480	1	US-08-282-581-4
c 26	30.6	13.7	480	1	US-08-550-544-4
27	30.6	13.7	1890	6	5312912-3

c 28	30.2	13.5	1368	3	US-08-874-563-5
c 29	30.2	13.5	1368	3	US-08-577-483-14
c 30	30.2	13.5	4253	3	US-08-577-483-7
c 31	30.2	13.5	7101	1	US-08-480-604A-9
c 32	30.2	13.5	7101	2	US-08-405-456A-9
c 33	30.2	13.5	7101	4	US-08-915-136-9
c 34	30.2	13.5	19124	2	US-08-487-826B-13
c 35	30.2	13.5	51952	3	US-08-947-823-1
c 36	30	13.5	638	4	US-08-991-789A-16
c 37	30	13.5	1140	4	US-08-893-654B-3
c 38	29.8	13.4	1104	4	US-09-277-716-30
c 39	29.8	13.4	43795	3	US-08-742-185-101
c 40	29.8	13.4	80246	4	US-09-078-294-4
c 41	29.6	13.3	870	2	US-08-420-629-9
c 42	29.6	13.3	1496	2	US-08-031-538-5
c 43	29.6	13.3	6152	4	US-08-973-462-1
c 44	29.4	13.2	732	4	US-08-916-576B-5
c 45	29.4	13.2	804	1	US-08-126-593A-3

ALIGNMENTS

RESULT 1

US-09-078-294-4  
; Sequence 4, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; CURRENT FILING DATE: 1998-05-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 80246  
; TYPE: DNA  
; ORGANISM: Nucleotide sequence of NC-contig  
US-09-078-294-4

Query Match 19.7%; Score 44; DB 4; Length 80246;

Best Local Similarity 58.3%; Pred No. 0.0015;

Matches 77; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 5 cataattataattataattataataactgaacacccatggtgccccctgctgttacatg 64

Db 15314 cataaatatacacataaatgtatatgttatgtatagacataaaatgtatatgtg 15373

Qy 65 gatttcctcgctactattgtatacgtgtatatataccgtataaagtacatatattata 124

Db 15374 tatatatacataaatgtatatgttatgtatatagacataaaatgtatatgtgata 15433

Qy 125 tgaacatgatta 136

Db 15434 tagacataaata 15445

RESULT 2

US-09-078-294-3  
; Sequence 3, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; CURRENT FILING DATE: 1998-05-13









;; MEDIUM TYPE: Disk, 3.5-inch  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: Windows NT  
;; SOFTWARE: Word97 & ASCII  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/058, 947A  
;; FILING DATE:  
;; CLASSIFICATION: 800  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 60/043,831  
;; FILING DATE: April 14, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: David J. Earp, Ph.D.  
;; REGISTRATION NUMBER: 41,401  
;; REFERENCE/DOCKET NUMBER: 5493-50032/DJE  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (503) 226-7391  
;; TELEFAX: (503) 228-9446  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3722  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; US-058-947A-1

Query Match 14.7%; Score 32.8; DB 4; Length 3722;  
Best Local Similarity 56.2%; Pred. No. 1.1;  
Matches 82; Conservative 0; Mismatches 62; Indels 2; Gaps 1;  
  
Qy 61 catggatttcctcactattgtatcacgtgtatatataccgtataatgtacatatatt 120  
Db 892 CGTAAATACAGAGAAATATAGGTACATATATATTCGGGTTTAAATGTCGATACATG 833  
  
Qy 121 tatatgaacatga--ttaatgcttctgtgagttgtctctcatccgttaagagtttcaatatgta 178  
Db 832 TATATCATCAAGAGTTACATGTTTTTGTATTTTACATTCGCAATTAATGTTCTTTTTCG 773  
  
Qy 179 atggtgaagagtcataaaccccaaatc 204  
Db 772 ATGCTTAATCGTCAACACCGATATC 747

RESULT 11  
US-08-947-823-1  
; Sequence 1, Application US/08947823  
; Patent No. 6114605  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Valerie M.  
; APPLICANT: Kaloshian, Isouhli  
; APPLICANT: Yaghoobi, Jafar  
; APPLICANT: Bodeau, John  
; APPLICANT: Milligan, Stephen  
; TITLE OF INVENTION: Procedures and Materials for Confering  
; TITLE OF INVENTION: Pest Resistance in Plants  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/947,823  
; FILING DATE: 09-OCT-1997  
; CLASSIFICATION: 800

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US97/18802  
;; FILING DATE: 09-OCT-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/028,191  
;; FILING DATE: 10-OCT-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bastian, Kevin L.  
;; REGISTRATION NUMBER: 34,774  
;; REFERENCE/DOCKET NUMBER: 023070-070210US  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 576-0200  
;; TELEFAX: (415) 576-0300  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 51952 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; US-08-947-823-1

Query Match 14.5%; Score 32.4; DB 3; Length 51952;  
Best Local Similarity 47.5%; Pred. No. 2.6;  
Matches 96; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
  
Qy 5 cataattataattataataactgaacacatggtgcctccctgctgttcatg 64  
Db 13606 CATCTACGTATCTATAAAATTCACATGAACACACGCTGTGAAGCGCATCTCAGAAA 13665  
  
Qy 65 gatttcctcctactattgtatcacgtgtatatataccgtataatgtacatatattata 124  
Db 13666 AACTAGTGTA 13725  
  
Qy 125 tgaacatgattaaagctgtgagttgtctcatccgttaagagtttcaatatgttaagtgtg 184  
Db 13726 TAT 13785  
  
Qy 185 aagagtcaaaccccaaatcat 206  
Db 13786 AATCTGGTTAAAAAAGATAAT 13807

RESULT 12  
US-08-443-639-7/c  
; Sequence 7, Application US/08443639  
; Patent No. 5981843  
; GENERAL INFORMATION:  
; APPLICANT: Chappell, Joseph  
; APPLICANT: Yin, Shaohui  
; APPLICANT: Cornett, Catherine A.G.  
; TITLE OF INVENTION: Transcriptional Control Sequences and  
; TITLE OF INVENTION: Methods  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/443,639  
; FILING DATE: 18-MAY-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33878  
REFERENCE/DOCKET NUMBER: 69-94  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4254 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Nicotiana tabacum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1217..1327, 1455..1718, 1806..2182,  
LOCATION: 2259  
LOCATION: ..2477, 2609..2747, 2903..3148, 3262..3558)  
US-08-443-639-7

Query Match 14.3%; Score 32; DB 2; Length 4254;  
Best Local Similarity 48.9%; Pred. No. 1.9;  
Matches 86; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 10 ttataattataattataactgaaccatggtgccccctgctgtttacatggattt 69  
DB 408 TTAATTTCAGAAAAAAGTTGCGTGGCGCGCGCATGTCACACACACAC 349  
QY 70 ctccgctactatttggatagctgtatataaccgtataatgacatatttatgaac 129  
DB 348 ATACATTAT 289  
QY 130 atgattaatgctgtgagttgtctccatccgtgaagcttccaatgtaagtga 185  
DB 288 ATATTAAATCTATAATTTTCTTAACTTTTCACAAATTTTAGTGTATTTTCTGA 233

RESULT 13  
US-08-446-855A-1/c  
Sequence 1, Application US/08446855A  
Patent No. 5849573  
GENERAL INFORMATION:  
APPLICANT: Stewart, Thomas S  
APPLICANT: Flores, Maria V  
APPLICANT: O'Sullivan, William J  
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl  
TITLE OF INVENTION: phosphate synthetase II  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderhye PC  
STREET: 1100 NO. 5849573th Glebe Road, 8th Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446.855A  
FILING DATE: 06-Jul-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29.009  
REFERENCE/DOCKET NUMBER: 47-80  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8920 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic  
US-08-446-855A-1

Query Match 14.3%; Score 32; DB 2; Length 8920;  
Best Local Similarity 54.2%; Pred. No. 2.2;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 77 actatttgatcgtgtatataaccgtataatgacatatttatgaacatgatta 136  
DB 8544 AATTTA 8485  
QY 137 atgcttgtagtggttctccatccgtgaagcttccaatgtaagtgaagagtcacaaac 196  
DB 8484 TACTTTTCATGTTGTTTGAATGAAAGTTAATAACATACATAAATAAGTAATGTTAAATC 8425

RESULT 14  
US-09-150-741-1/c  
Sequence 1, Application US/09150741  
Patent No. 6183996  
GENERAL INFORMATION:  
APPLICANT: Stewart et al.  
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate  
Patent No. 6183996  
TITLE OF INVENTION: Synthetase II  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/09/150,741  
CURRENT FILING DATE: 1998-09-10  
EARLIER APPLICATION NUMBER: PL6380  
EARLIER FILING DATE: 1992-12-16  
EARLIER APPLICATION NUMBER: AU93/00617  
EARLIER FILING DATE: 1993-12-02  
EARLIER APPLICATION NUMBER: 08/446.855  
EARLIER FILING DATE: 1995-07-06  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 8920  
TYPE: DNA  
ORGANISM: Plasmodium falciparum  
US-09-150-741-1

Query Match 14.3%; Score 32; DB 4; Length 8920;  
Best Local Similarity 54.2%; Pred. No. 2.2;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 77 actatttgatcgtgtatataaccgtataatgacatatttatgaacatgatta 136  
DB 8544 AATTTA 8485  
QY 137 atgcttgtagtggttctccatccgtgaagcttccaatgtaagtgaagagtcacaaac 196  
DB 8484 TACTTTTCATGTTGTTTGAATGAAAGTTAATAACATACATAAATAAGTAATGTTAAATC 8425

RESULT 15  
US-08-282-581-4  
Sequence 4, Application US/08282581  
Patent No. 5670349  
GENERAL INFORMATION:  
APPLICANT: Cramer, Carole L.  
APPLICANT: Weissenborn, Deborah L.  
TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:10 ; Search time 8261.74 Seconds  
(without alignments)  
290.049 Million cell updates/sec

Title: US-09-698-903B-10\_COPY\_194\_416  
Perfect score: 223  
Sequence: 1 ctaccataattataattata.....catgaacacccaactcgat 223

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
C 1	45.6	20.4	AA605352 30388 Lam
C 2	44.4	19.9	AU088482
C 3	43.6	19.6	BE948306
C 4	42	18.8	1024 13 CNS030D0
C 5	41.8	18.7	487 10 AU087926
C 6	41.8	18.7	500 10 AU088152
C 7	41.6	18.7	527 13 AQ172252
C 8	41.6	18.7	599 13 AQ273686
C 9	41.6	18.7	711 13 AZ119444
C 10	41.6	18.7	789 13 AZ126726
C 11	40.5	18.2	571 13 AQ797772
C 12	40.6	18.2	605 13 AZ996260

C 13	40.4	18.1	583 13 AQ774291
C 14	40.4	18.1	595 10 AA394343
C 15	40.4	18.1	845 13 AQ745537
C 16	40.2	18.0	525 11 BF460992
C 17	39.8	17.8	612 13 AZ409529
C 18	39.8	17.8	836 13 AZ908564
C 19	39.6	17.8	456 11 BG836222
C 20	39.4	17.7	458 13 BH040914
C 21	39.4	17.7	765 13 AZ705201
C 22	39.4	17.7	939 13 CNS0403P
C 23	39.2	17.6	363 13 AZ598385
C 24	39.2	17.6	1020 13 CNS03AUS
C 25	39.2	17.6	1101 13 CNS00DGO
C 26	39	17.5	309 13 AQ542682
C 27	39	17.5	511 11 BG633595
C 28	39	17.5	610 11 BF296565
C 29	39	17.5	764 13 AQ915359
C 30	39	17.5	834 10 AW941790
C 31	38.6	17.3	374 13 BH126294
C 32	38.6	17.3	453 13 AQ818698
C 33	38.6	17.3	596 13 BH126267
C 34	38.4	17.2	500 10 AU087788
C 35	38.4	17.2	536 13 AZ013131
C 36	38.4	17.2	563 13 AQ561057
C 37	38.2	17.1	192 10 AA547848
C 38	38.2	17.1	543 13 AZ060650
C 39	38.2	17.1	604 13 AZ347990
C 40	38	17.0	514 13 AQ141523
C 41	38	17.0	541 13 AZ640886
C 42	38	17.0	609 13 AQ367613
C 43	37.8	17.0	324 13 BH126635
C 44	37.8	17.0	412 10 A1991410
C 45	37.8	17.0	416 13 AZ869448

ALIGNMENTS

RESULT 1  
AA605352/c  
LOCUS 30388 Lambda-PRL2 Arabidopsis thaliana cDNA clone 269H8T7, mRNA  
DEFINITION sequence.  
ACCESSION AA605352  
VERSION AA605352.1 GI:2445880  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
REFERENCE 1 (bases 1 to 507)  
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel,E. and Somerville,C.  
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones Plant Physiol. 106, 1241-1255 (1994)  
JOURNAL 95148729  
MEDLINE  
COMMENT Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9168  
Email: 22313tcn@bm.cl.msu.edu  
Seq primer: T7 dye primer.  
LOCATION/Qualifiers  
1. 507  
/organism="Arabidopsis thaliana"  
/strain="var columbia"  
/db\_xref="taxon:3702"

Query Match  
 Best Local Similarity 20.4%; Score 45.6; DB 10; Length 507;  
 Matches 75; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 Qy 93 tatataccgtataatgacatatttatataacatgatttaagtctgtgagttgtt 152  
 Db 336 TATATATATATGTCATATGCCCCAGTATGTATATAACATGACAACTCAATGCTGTGTTTC 277  
 Qy 153 ctcatccgtgaagcttccaatgtaagtgtgagagtcacaaaccccaaatcatgaacac 212  
 Db 276 ATCTCCATAGAGTTGCCAATATGTCATGTTGACAGTCAAAACCCCAAAATCATGAACAC 217  
 Qy 213 ccaa 216  
 Db 216 CAAA 213

BASE COUNT 138 a 94 c 112 g 144 t 19 others  
 ORIGIN

/clone="269H8T7"  
 /note="Vector: lambda zip-Lox; Site\_1: Sal; Site\_2: Not;  
 Lambda PRL2 is a cDNA library derived from equal  
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
 day germinated etiolated seedlings; 2) tissue culture  
 grown roots; 3) staged plants half with 24 hour light  
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
 same plants as 3 but aerial tissue (stems, flowers and  
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
 inserts were directionally cloned with Sal-Not arms using  
 oligo dr primed cDNA. "

Query Match  
 Best Local Similarity 20.4%; Score 45.6; DB 10; Length 507;  
 Matches 75; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 Qy 93 tatataccgtataatgacatatttatataacatgatttaagtctgtgagttgtt 152  
 Db 336 TATATATATATGTCATATGCCCCAGTATGTATATAACATGACAACTCAATGCTGTGTTTC 277  
 Qy 153 ctcatccgtgaagcttccaatgtaagtgtgagagtcacaaaccccaaatcatgaacac 212  
 Db 276 ATCTCCATAGAGTTGCCAATATGTCATGTTGACAGTCAAAACCCCAAAATCATGAACAC 217  
 Qy 213 ccaa 216  
 Db 216 CAAA 213

BASE COUNT 138 a 94 c 112 g 144 t 19 others  
 ORIGIN

/clone="269H8T7"  
 /note="Vector: lambda zip-Lox; Site\_1: Sal; Site\_2: Not;  
 Lambda PRL2 is a cDNA library derived from equal  
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
 day germinated etiolated seedlings; 2) tissue culture  
 grown roots; 3) staged plants half with 24 hour light  
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
 same plants as 3 but aerial tissue (stems, flowers and  
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
 inserts were directionally cloned with Sal-Not arms using  
 oligo dr primed cDNA. "

Query Match  
 Best Local Similarity 20.4%; Score 45.6; DB 10; Length 507;  
 Matches 75; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 Qy 93 tatataccgtataatgacatatttatataacatgatttaagtctgtgagttgtt 152  
 Db 336 TATATATATATGTCATATGCCCCAGTATGTATATAACATGACAACTCAATGCTGTGTTTC 277  
 Qy 153 ctcatccgtgaagcttccaatgtaagtgtgagagtcacaaaccccaaatcatgaacac 212  
 Db 276 ATCTCCATAGAGTTGCCAATATGTCATGTTGACAGTCAAAACCCCAAAATCATGAACAC 217  
 Qy 213 ccaa 216  
 Db 216 CAAA 213

BASE COUNT 138 a 94 c 112 g 144 t 19 others  
 ORIGIN

/clone="269H8T7"  
 /note="Vector: lambda zip-Lox; Site\_1: Sal; Site\_2: Not;  
 Lambda PRL2 is a cDNA library derived from equal  
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
 day germinated etiolated seedlings; 2) tissue culture  
 grown roots; 3) staged plants half with 24 hour light  
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
 same plants as 3 but aerial tissue (stems, flowers and  
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
 inserts were directionally cloned with Sal-Not arms using  
 oligo dr primed cDNA. "

Query Match  
 Best Local Similarity 20.4%; Score 45.6; DB 10; Length 507;  
 Matches 75; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 Qy 93 tatataccgtataatgacatatttatataacatgatttaagtctgtgagttgtt 152  
 Db 336 TATATATATATGTCATATGCCCCAGTATGTATATAACATGACAACTCAATGCTGTGTTTC 277  
 Qy 153 ctcatccgtgaagcttccaatgtaagtgtgagagtcacaaaccccaaatcatgaacac 212  
 Db 276 ATCTCCATAGAGTTGCCAATATGTCATGTTGACAGTCAAAACCCCAAAATCATGAACAC 217  
 Qy 213 ccaa 216  
 Db 216 CAAA 213

BASE COUNT 138 a 94 c 112 g 144 t 19 others  
 ORIGIN

/clone="269H8T7"  
 /note="Vector: lambda zip-Lox; Site\_1: Sal; Site\_2: Not;  
 Lambda PRL2 is a cDNA library derived from equal  
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
 day germinated etiolated seedlings; 2) tissue culture  
 grown roots; 3) staged plants half with 24 hour light  
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
 same plants as 3 but aerial tissue (stems, flowers and  
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
 inserts were directionally cloned with Sal-Not arms using  
 oligo dr primed cDNA. "

Query Match  
 Best Local Similarity 20.4%; Score 45.6; DB 10; Length 507;  
 Matches 75; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 Qy 93 tatataccgtataatgacatatttatataacatgatttaagtctgtgagttgtt 152  
 Db 336 TATATATATATGTCATATGCCCCAGTATGTATATAACATGACAACTCAATGCTGTGTTTC 277  
 Qy 153 ctcatccgtgaagcttccaatgtaagtgtgagagtcacaaaccccaaatcatgaacac 212  
 Db 276 ATCTCCATAGAGTTGCCAATATGTCATGTTGACAGTCAAAACCCCAAAATCATGAACAC 217  
 Qy 213 ccaa 216  
 Db 216 CAAA 213

BASE COUNT 138 a 94 c 112 g 144 t 19 others  
 ORIGIN

/clone="269H8T7"  
 /note="Vector: lambda zip-Lox; Site\_1: Sal; Site\_2: Not;  
 Lambda PRL2 is a cDNA library derived from equal  
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
 day germinated etiolated seedlings; 2) tissue culture  
 grown roots; 3) staged plants half with 24 hour light  
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
 same plants as 3 but aerial tissue (stems, flowers and  
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
 inserts were directionally cloned with Sal-Not arms using  
 oligo dr primed cDNA. "

Query Match  
 Best Local Similarity 20.4%; Score 45.6; DB 10; Length 507;  
 Matches 75; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 Qy 93 tatataccgtataatgacatatttatataacatgatttaagtctgtgagttgtt 152  
 Db 336 TATATATATATGTCATATGCCCCAGTATGTATATAACATGACAACTCAATGCTGTGTTTC 277  
 Qy 153 ctcatccgtgaagcttccaatgtaagtgtgagagtcacaaaccccaaatcatgaacac 212  
 Db 276 ATCTCCATAGAGTTGCCAATATGTCATGTTGACAGTCAAAACCCCAAAATCATGAACAC 217  
 Qy 213 ccaa 216  
 Db 216 CAAA 213

BASE COUNT 138 a 94 c 112 g 144 t 19 others  
 ORIGIN

/clone="269H8T7"  
 /note="Vector: lambda zip-Lox; Site\_1: Sal; Site\_2: Not;  
 Lambda PRL2 is a cDNA library derived from equal  
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7  
 day germinated etiolated seedlings; 2) tissue culture  
 grown roots; 3) staged plants half with 24 hour light  
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
 same plants as 3 but aerial tissue (stems, flowers and  
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA  
 inserts were directionally cloned with Sal-Not arms using  
 oligo dr primed cDNA. "

Best Local Similarity 53.2%; Pred. No. 0.44;  
 Matches 115; Conservative 0; Mismatches 98; Indels 3; Gaps 1;  
 Qy 2 tacataataattataataattataataactgaaccatgggtgccccctgctgctttac 61  
 Db 240 TACAAT 181  
 Qy 62 atgagttctccgctactatttgcacgtgtatataaccgtata---atgtacatatata 118  
 Db 180 TTTTATAGAAATGATA 121  
 Qy 119 ttatatgaacatgatttaagtgtgagttgttctcatccgtgaagagtttccaatatgta 178  
 Db 120 TTTTAAAAAATTTTAAATTTTATGTTGTTTTCGTACCTATATATCTATAAATATAT 61  
 Qy 179 atggtgaagagtcacaaaccccaaatcatgaacacc 214  
 Db 60 TTCAATAGATATATAACNCAANATCATATATATCC 25

RESULT 3  
 BE948306  
 LOCUS  
 DEFINITION UI-M-BH3-awp-b-11-0-UI.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
 ACCESSION BE948306  
 VERSION UI-M-BH3-awp-b-11-0-UI 3', mRNA sequence.  
 KEYWORDS BE948306.1 GI:10526065  
 SOURCE EST.  
 ORGANISM house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 465)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: MEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to identify it as a clone from the  
 cerebellum tissue cDNA Library Preparation: M.B. Soares Lab Clone  
 distribution: Researchers may obtain BMAP cDNA clones from RESEARCH  
 GENETICS. It should be noted that Bento Soares is generating a  
 small number of additional specialized non-redundant arrays of BMAP  
 cDNAs whose availability will be considered under appropriate and  
 limited collaborative arrangements. The following repetitive  
 elements were found in this cDNA sequence: 255-326, >Lx9LINE/11  
 Seg primer: M13 Forward  
 POLYA=Yes

FEATURES  
 Location/Qualifiers  
 1..465  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-awp-b-11-0-UI"  
 /clone\_lib="NIH\_BMAP\_M\_S4"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged  
 normalized libraries from ten regions of the mouse brain





[illegible]



ACCESSION	AZ126726
VERSION	AZ126726.1 GI:8201718
KEYWORDS	GSS.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriharoidae; Oryzede; Oryza. 1 (bases 1 to 789) Wing,R.A. and Dean,R.A. A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998) Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: GTAAACGACGCCAGTG Class: BAC ends High quality sequence start: 21 High quality sequence stop: 683. Location/Qualifiers 1..789 /organism="Oryza sativa" /strain="Japonica" /cultivar="Nipponbare" /db_xref="taxon:4530" /clone="OSJNB0079C19f" /clone.lib="CUGI Rice BAC Library (ECORI)" /tissue_type="Leaf" /lab_host="E. coli DH10B" /note=vector: pACindigo; Site_1: EcoRI; Site_2: EcoRI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ("www.genome.clemson.edu")."
BASE COUNT	292 a 124 c 111 g 253 t
ORIGIN	9 others
Query Match	18.7%; Score 41.6; DB 13; Length 789;
Best Local Similarity	50.9%; Pred. No. 2,1;
Matches	86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY	9 attataattataatataactgaaccctgggtcccccgtgctttacatgatt 68 
Db	745 ATTATATATTATTTATATAAAAATATATATTTTATATCTCTTTTANNTNN 686
QY	69 tcctcgctactatttgtaacgttgttatataccgtataatgtcacatatattatga 128 
Db	685 NATTANNNTATTTATATATATATATATATATATATATATATATATATATA 626
QY	129 catgatgaatgcttgtagttgttcctaccctaagaggtttcaaatgt 177 
Db	625 TATATATCTTCATGCAGATGTTCACAPACAGATAGCTTTTAGGTAT 577



**ORGANISM** Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 595)  
**REFERENCE** Newman,T., deBrujin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh  
L., Ohlrogger,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel  
,E. and Somerville,C.  
**TITLE** Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
**JOURNAL** Plant Physiol. 106, 1241-1255 (1994)  
**MEDLINE** 95148729  
**COMMENT** Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRU, Michigan State University,Plant Biology Bldg.,E.  
Lansing,MI  
Tel.: 517-353-0854  
Fax: 517-353-9168  
Email: 22313t@bm.cl.msu.edu  
Seq primer: 77.

**FEATURES**

source	Location/Qualifiers
	1..595
	/organism="Arabidopsis thaliana"
	/strain="var columbia"
	/db_xref="taxon:3702"
	/clone="310A7T7"
	/clone_lib="Lambda-PRL2"
	/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
	Lambda PRL2 is a cDNA library derived from equal
	quantities of 4 pools of mRNA. The mRNA sources were 1) 7
	day germinated etiolated seedlings; 2) tissue culture
	grown roots; 3) staged plants half with 24 hour light
	cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
	same plants as 3 but aerial tissue (stems, flowers and
	siliques. The vector is BRL's lambda Zip-Lox. The cDNA
	inserts were directionally cloned with Sal-Not arms using
	oligo dt primed cDNA."

175 a 107 c 130 g 172 t 11 others

**BASE COORDINATES**

**ORIGIN**

Query Match	18.1%; Score 40.4; DB 10; Length 595;
Best Local Similarity	63.3%; Pred. No. 4.1;
Matches	62; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```

QY      119 ttatatgaacatgattaatgcctgttgagttgttcctcgcgaagagtccaatgatga 178
Ddb     ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      310 TATGTTAAACATGAACAACACTCAATCGTGTTTCATCCATAGAGTTGCCAATATGTCA 251
Ddb     ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      179 atggtagagatccaaaccacaaaatcatgaacacccc 216
Ddb     ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
QY      250 ATGGTGAACAGTCAAAAACCAAAATCATGAACACCAA 213
Ddb     ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||

```

**RESULT 15**

**LOCUS** AQ745537/c

**DEFINITION** HS\_2272\_A2\_B12\_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2272 Col=24 Row=C, DNA sequence.

**ACCESSION** AQ745537

**VERSION** AQ745537.1 GI:5523059

**KEYWORDS** GSS.

**SOURCE** human.

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 845)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,K., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:02:29 ; Search time 2331.3 Seconds  
(without alignments)  
162.757 Million cell updates/sec

Title: us-09-698-903B-11

Perfect score: 23  
Sequence: 1 gaaatccatgtaaacgacgagg 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_em.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB ID	Description
1	23	100.0	23	6	AX127758 Sequence
2	23	100.0	415	6	AX127755 Sequence
3	23	100.0	416	6	AX127757 Sequence
4	18.8	81.7	41760	2	AC090679 Homo sapi
5	18.8	81.7	89350	8	AF112H17 Arabidops
6	18.8	81.7	156515	2	AC022947 Homo sapi
7	18.8	81.7	199577	8	ATCHRIV57 Arabidops
8	18.8	81.7	208684	2	AC072048 Mus muscu
9	18.4	80.0	188524	2	AC046148 Mus muscu
10	18.4	80.0	172650	2	AC040920 Homo sapi
11	18.2	79.1	535	5	AF118429 Falco per
12	18.2	79.1	146285	9	AC005083 Homo sapi
13	18.2	79.1	183634	2	AC023167 Mus muscu
14	18.2	79.1	228809	2	AC021443 Homo sapi
15	17.8	77.4	23486	8	AC007294 Arabidops
16	17.8	77.4	26275	3	U29377 Caenorhabdi
17	17.8	77.4	94835	9	HS0316D7 Human DNA
18	17.8	77.4	139959	33	AC013597 Homo sapi
19	17.8	77.4	146746	9	HS03222 Human DNA s
20	17.8	77.4	149755	2	AC092795 Homo sapi
21	17.8	77.4	168749	2	AC008573 Homo sapi
22	17.8	77.4	177703	9	AC019046 Homo sapi
23	17.8	77.4	179665	2	AC010246 Homo sapi
24	17.8	77.4	188537	2	AC090071 Homo sapi
25	17.8	77.4	213890	2	AC013461 Homo sapi
26	17.8	77.4	253038	2	AC008930 Homo sapi
27	17.4	75.7	25042	9	HS380A1 Human DNA s
28	17.4	75.7	89939	2	AC022667 Homo sapi
29	17.4	75.7	106392	9	AC007316 Homo sapi
30	17.4	75.7	113033	9	AC005070 Homo sapi
31	17.4	75.7	118904	9	HS238G2 Human DNA
32	17.4	75.7	134506	9	HS357I16 Homo sapi
33	17.4	75.7	136497	2	HS03448H8 Homo sapi
34	17.4	75.7	155316	2	AC078902 Homo sapi
35	17.4	75.7	155932	1	CJ11168X6 Campyloba
36	17.4	75.7	156014	5	AF003796 Gallus ga
37	17.4	75.7	171368	9	HSU95738 Human chrom
38	17.4	75.7	171868	9	AL359697 Human DNA
39	17.4	75.7	194873	2	AL592440 Homo sapi
40	17.4	75.7	256073	9	AE006464 Homo sapi
41	17.2	74.8	466	11	HSPE53B06 H. sapiens
42	17.2	74.8	522	5	RCABCRYST X87114 R. catesbeia
43	17.2	74.8	985	3	AB002187 Epiliachna
44	17.2	74.8	985	3	AB002188 Epiliachna
45	17.2	74.8	985	3	AB002189 Epiliachna

ALIGNMENTS

RESULT 1

AX127758 Locus

AX127758 Definition Sequence 11 from Patent WO0131042.

AX127758 Accession

AX127758 Version

AX127758.1 Keywords

GI:14134405

ORGANISM

synthetic construct.

synthetic construct

artificial sequence.

1 (bases 1 to 23)

REFERENCE

WESTON, B. and de Beuckeleer, M.

AUTHORS

Male-sterile brassica plants and methods for producing same

TITLE

Patent: WO 0131042-A 11 03-MAY-2001;

JOURNAL

Aventis CropScience N.V. (BE)

FEATURES

location/Qualifiers

1..23

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer MDB371"

PAT 15-MAY-2001

BASE COUNT 9 a 4 c 7 g 3 t  
ORIGIN

Query Match 100.0%; Score 23; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.26;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgacgagg 23  
|||||  
Db 1 GAAATCCATGTAAGCAGCAGGG 23

## RESULT 2

AX127755  
LOCUS AX127755 415 bp DNA PAT 15-MAY-2001  
DEFINITION Sequence 8 from Patent WO0131042.  
ACCESSION AX127755  
VERSION AX127755.1 GI:14134402  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 415).  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 8 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source 1. .415  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="5' border flanking region of elite event MS-B2"  
misc\_feature 1. .234  
/note="plant DNA"  
misc\_feature 235. .415  
/note="T-DNA"

BASE COUNT 154 a 55 c 70 g 136 t  
ORIGIN

Query Match 100.0%; Score 23; DB 6; Length 415;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgacgagg 23  
|||||  
Db 152 GAAATCCATGTAAGCAGCAGGG 174

## RESULT 3

AX127757/c  
LOCUS AX127757 416 bp DNA PAT 15-MAY-2001  
DEFINITION Sequence 10 from Patent WO0131042.  
ACCESSION AX127757  
VERSION AX127757.1 GI:14134404  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 416).  
AUTHORS Weston,B. and de Beuckeleer,M.  
TITLE Male-sterile brassica plants and methods for producing same  
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source 1. .416  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="5' border flanking region of elite event MS-B2"  
misc\_feature 1. .193  
/note="T-DNA"  
misc\_feature 194. .416

BASE COUNT 137 a 72 c 54 g 152 t 1 others  
ORIGIN

Query Match 100.0%; Score 23; DB 6; Length 416;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgacgagg 23  
|||||  
Db 263 GAAATCCATGTAAGCAGCAGGG 241

## RESULT 4

AC090679/c  
LOCUS AC090679 41760 bp DNA HTG 27-JUN-2001  
DEFINITION Homo sapiens chromosome 12q clone RP11-8003, \*\*\* SEQUENCING IN PROGRESS \*\*\*; 14 unordered pieces.  
AC090679  
VERSION AC090679.3 GI:14190581  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 41760)

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarella,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,D., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwkw,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 41760)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One





```

/exon      /gene="T12H17.20"
/number=1
2844..2983
/exon      /gene="T12H17.20"
/number=2
3327..3400
/note="Asn-tRNA_(GTT)"
4696..5489
/exon      /gene="T12H17.30"
join(4696..5035,5407..5489)
/exon      /gene="T12H17.30"
/note="contains EST gb:N38564"
/codon_start=1
/product="putative protein"
/protein_id="CAA16549.1"
/db_xref="GI:2827541"
/db_xref="SPTREMBL:O49645"
/translation="MAYTNKVTISAATVATMMLFLAVTIVDAQSMPPMPKPNPVCALAD
LPNIQICYFNLDITPSECCNDLKSSSTIQVNCLCDNFIAHPNSGNISQARYDLVNS
ACGVADFEACKGDSAGGSTNKIAASVLLGLVASLFF"
/exon      /gene="T12H17.30"
/number=1
5036..5406
/exon      /gene="T12H17.30"
/number=1
5407..5489
/exon      /gene="T12H17.30"
/number=2
join(6595..6940,7035..7105)
/exon      /gene="T12H17.40"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAA16550.1"
/db_xref="GI:2827542"
/db_xref="SPTREMBL:O49646"
/translation="MAVAVTAVLFLAVIAPQWETKPPRESDTSDTSGTSGDRDRRT
MCLPSLPGVQNCYATFLNAPPSKECKDLKTASKREVTCLCNNVIAHPDPLTYNTNQV
6595..6940
/exon      /gene="T12H17.40"
/number=1
6595..7105
/exon      /gene="T12H17.40"
6941..7034
/exon      /gene="T12H17.40"
/number=1
7035..7105
/exon      /gene="T12H17.40"
/number=2
complement(8019..9209)
/exon      /gene="T12H17.50"
/number=1
8019..9209
/exon      /gene="T12H17.50"
complement(8019..9209)
/exon      /gene="T12H17.50"
complement(8019..9209)
/exon      /gene="T12H17.50"
/note="similarity to predicted protein, Arabidopsis
thaliana"
/codon_start=1
/product="predicted protein"
/protein_id="CAA16551.1"
/db_xref="GI:2827543"
/db_xref="SPTREMBL:O49647"
/translation="MEKNQNPNTWSDLPDLILNLVFKRLSFANFRQAKSCSSWYSAS
KQSPKQIPIWMLFPKDKNNKNSCTIFNPEDKQLYQTODLGVFEAKSVCLATY
GSLWLDQDSKYNLILNPFYETIGLPAIESQOVGMKVQDTIDDDPLTDDHNVKL
FKGNVTVPVWIDEKTEYIALWGLGYWCVVYAKNGDKLWQIPEIILDSIDMWYK
DHKLYSFSLNFTILDFSGEIPKRAFCQCFMHVYRSPLSPRSQIQLNSWCVAETKL
VTVTGDLVLLVHWSRIQSENKYFSSGFLDKYELASIGDEAMLLDGLTVL
ANEVEGLHRTIYFSDSHDITTKDLFLFNTRYETREMEPLHKFDKCSLLELSARWFLPS
FSHT"
9990..13709
/exon      /gene="T12H17.60"
join(9990..10023,10239..10535,10564..10664,10910..11292,
11755..11809,11898..12053,12143..12311,12391..12455,
12534..12674,12769..12885,13024..13096,13244..13380,
13468..13525,13612..13709)
/exon      /gene="T12H17.60"
/note="similarity to p48, Homo sapiens, PATCHX:G904032
contains EST gb:Ag605571, H76919, H76638, R30271, Z25650"
/codon_start=1
/product="HSP associated protein like"
/protein_id="CAA16552.1"
/db_xref="GI:2827544"
/db_xref="SPTREMBL:O49648"
/translation="HRTWVLAESVVAIAPLLAPQSMFPMTEPECATVMPDLLEK
CPATGSVTPECCDLKTSATQVTCLDNYIANPNSITGPYSKAITTKGVFQK
YSCDGTGSKVSNKNGAFMVFINGFWCVKKEKKKEAAVAMGVTLGLINGLIVAH
GKTRQONKPVCKLILFQFHQFCISIRFSANKKRSKRTGDAYPFRNEKVLQKPYS
LPFVQILEIETRLKTKKTRLLTRKPTLFLFLSLSRSEADFQKQGFNEAYLGAK
IPTGVHEEDTKPRSFVVEESDDMDTEEVKPKVEEEDEDEIVESDVELEGTVE
PNDPPQKMGDSSEVTDENREAAQEAQKAMEALSEGFEAIEHLTRAITLNPISA
INYNRASVYIKLKPNAIRDANAALINPDASKYKSRGMARMLGEAAAKDLH
LASTIDYDEEISAVLKKVEPNAHLEEHRRKYDRLKEREDKKAERDLRRRAQAA
YDKAKKEQSSSRSPSGGFGPGMGGMGPAGMGGMGPAGMGGMGPAGMGGMGP
GGMDFSKILNDPELMTAFSDPEVMAALQDVNMKNPANLAKHQANPKVAPVIAKMGKFA
GPQ"
9990..10023
/exon      /gene="T12H17.60"
/number=1
10024..10238
/exon      /gene="T12H17.60"
/number=1
10239..10535
/exon      /gene="T12H17.60"
/number=2
10536..10563
/exon      /gene="T12H17.60"
/number=2
10564..10664
/exon      /gene="T12H17.60"
/number=3
10665..10909
/exon      /gene="T12H17.60"
/number=3
10910..11292
/exon      /gene="T12H17.60"
/number=4
11293..11754
/exon      /gene="T12H17.60"
/number=4
11755..11809
/exon      /gene="T12H17.60"
/number=5
11810..11897
/exon      /gene="T12H17.60"
/number=5
11898..12053
/exon      /gene="T12H17.60"
/number=6
12054..12142
Query Match      81.7%  Score 18.8;  DB 8;  Length 89350;
Best Local Similarity 90.9%;  Pred. No. 95;
Matches 20;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;
QY 2 aaatccatgtaaagcagcaggg 23
|||||
Db 15674 AAATCCATGTAAAGCAGCTAAGG 15653
RESULT 6
AC022947/c

```

```

LOCUS       AC022947       156515 bp       DNA             HTG             12-MAR-2000
DEFINITION   Homo sapiens clone RP11-8003, WORKING DRAFT SEQUENCE, 13 unordered
              pieces.
ACCESSION    AC022947
VERSION      AC022947.2   GI:7229879
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 156515)
BIRREN,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-8003
Unpublished
              2 (bases 1 to 156515)
BIRREN,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
Choelel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivier,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6921807.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6397
Center Clone name: 80_O_3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147422 bases at least Q40
Consensus quality: 151928 bases at least Q30
Consensus quality: 153939 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 153115; sum-of-contigs
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
              1 1026: contig of 1026 bp in length
              * 1027 1126: gap of 100 bp
              * 1127 5787: contig of 4661 bp in length
              * 5788 5887: gap of 100 bp
              * 5888 11595: contig of 5708 bp in length
              * 11596 11695: gap of 100 bp

```

```

* 11696 14431: contig of 2736 bp in length.
* 14432 14531: gap of 100 bp
* 14532 20088: contig of 5557 bp in length
* 20089 20188: gap of 100 bp
* 20189 29619: contig of 9431 bp in length
* 29620 29719: gap of 100 bp
* 29720 38087: contig of 8368 bp in length
* 38088 38187: gap of 100 bp
* 38188 50251: contig of 12064 bp in length
* 50252 50351: gap of 100 bp
* 50352 60663: contig of 10312 bp in length
* 60664 60763: gap of 100 bp
* 60764 75537: contig of 14774 bp in length
* 75538 75637: gap of 100 bp
* 75638 93139: contig of 17502 bp in length
* 93140 93239: gap of 100 bp
* 93240 114372: contig of 21133 bp in length
* 114373 114472: gap of 100 bp
* 114473 156515: contig of 42043 bp in length.
FEATURES             Location/Qualifiers
     source            1..156515
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone_lib="RPC1-11 Human Male BAC"
     misc_feature      1..1026
                       /note="assembly_fragment"
     misc_feature      1127..5787
                       /note="assembly_fragment"
     misc_feature      5888..11595
                       /note="assembly_fragment"
     misc_feature      11696..14431
                       /note="assembly_fragment"
                       /clone_end:SP6
                       /vector_side:left
     misc_feature      14532..20088
                       /note="assembly_fragment"
     misc_feature      20189..29619
                       /note="assembly_fragment"
                       /clone_end:T7
     misc_feature      29720..38087
                       /note="assembly_fragment"
     misc_feature      38188..50251
                       /note="assembly_fragment"
     misc_feature      50352..60663
                       /note="assembly_fragment"
     misc_feature      60764..75537
                       /note="assembly_fragment"
     misc_feature      75638..93139
                       /note="assembly_fragment"
     misc_feature      93240..114372
                       /note="assembly_fragment"
     misc_feature      114473..156515
                       /note="assembly_fragment"
BASE COUNT  49638 a 27072 c 27170 g 51430 t 1205 others
ORIGIN

```

Query Match 81.7%; Score 18.8; DB 2; Length 156515;  
 Best Local Similarity 90.9%; Pred. No. 1e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaaatcatgtaaagcagcagg 22  
 ||| ||||| ||||| |||||

Db 76912 GAATTCATGTAAAGCTGCAGG 76891

RESULT 7

ATCHRIV57/c  
 LOCUS ATCHRIV57 199577 bp DNA 16-MAR-2000  
 DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 57.  
 ACCESSION AL161557

VERSION	AL161557.2	GI:7269071	
KEYWORDS	thale cress.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.		
REFERENCE	1 (bases 64578 to 155510)		
JOURNAL	Wedler,H., Wambutt,R., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
REFERENCE	Unpublished		
AUTHORS	2 (bases 141638 to 141797)		
JOURNAL	Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
REFERENCE	Unpublished		
AUTHORS	3 (bases 143670 to 199577)		
JOURNAL	Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
REFERENCE	Unpublished		
AUTHORS	4 (bases 48809 to 64829)		
JOURNAL	Robben,J., Grymonprez,B., Volckaert,G, Mewes,H.W., Lemcke,K. and Mayer,K.F.X.		
REFERENCE	Unpublished		
AUTHORS	5 (bases 1 to 199577)		
JOURNAL	EU Arabidopsis sequencing,project.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de		
TITLE	Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk		
JOURNAL	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/		
REFERENCE	this fragment has an overlap with ATCHRIV56 at the 5' end and an overlap with ATCHRIV58 at the 3' end.		
COMMENT	Location/Qualifiers		
FEATURES	1..199577		
source	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="4"		
	6674..9057		
gene	/gene="AT4g22260"		
exon	complement(6674..6882)		
	/gene="AT4g22260"		
	/number=1		
gene	complement(join(6674..6882,7268..7339,7507..7575,7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))		
	/gene="AT4g22260"		
CDS	complement(join(6674..6882,7268..7339,7507..7575,7667..7775,7863..7978,8361..8454,8588..8777,8909..9057))		
	/gene="AT4g22260"		
	/note="similarity to alternative oxidase, Mangifera indica, PIR2:S45035		
	contains EST gb:T42793, AI994896.1, AI997645.1, R30022, AA395166"		
	/codon_start=1		
	/product="putative protein"		
	/protein_id="CAB79181.1"		
	/db_xref="GI:7269072"		
	/translation="MAAISGISGTLTISRLPLTLRRRAAVSYSSSHRLHLPLSS RRLLNRRHQATILQDDEKVVVEESFKAETGTGTEPLEPNSSSTSAFETWIL KLEQVNVFLTQDSDIKIDLTLYRDTVARFVLETIARVPVFAFMSVLHMYETGWR RADYLVKHFAESWNEMHLLIMEELGGNSWDFRLAQHIATFYFMVFIYLISPRM AVHFSCEVSHAYETDQKFLKASGEELKNMPADIAVKYTGGLYLFLNLYDFVNI RDEAECHKTHRACTGLSLRSPHLEDDDDTEESGCVVPEAHCEGIVDCLKKSI S"		
intron	complement(6883..7267)		
	/number=1		
exon	complement(7268..7339)		
	/gene="AT4g22260"		
	/number=2		
	complement(7340..7506)		
	/number=2		
	/gene="AT4g22260"		
	/number=3		
intron	complement(7507..7575)		
	/number=3		
exon	complement(7576..7666)		
	/number=3		
intron	complement(7667..7775)		
	/gene="AT4g22260"		
	/number=4		
intron	complement(7776..7862)		
	/number=4		
exon	complement(7863..7978)		
	/gene="AT4g22260"		
	/number=5		
intron	complement(7979..8360)		
	/number=5		
exon	complement(8361..8454)		
	/gene="AT4g22260"		
	/number=6		
intron	complement(8455..8587)		
	/number=6		
exon	complement(8588..8777)		
	/gene="AT4g22260"		
	/number=7		
intron	complement(8778..8908)		
	/number=7		
exon	complement(8909..9057)		
	/gene="AT4g22260"		
	/number=8		
exon	10103..10244		
	/gene="AT4g22270"		
	/number=1		
gene	10103..12393		
	/gene="AT4g22270"		
CDS	join(10103..10244,10395..10516,11327..11779,11868..12174,12254..12393)		
	/gene="AT4g22270"		
	/note="contains EST gb:AI997575.1"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="CAB79182.1"		
	/db_xref="GI:7269073"		
	/translation="MATNGSSEHLLPETTPRSALINKQOPEEPASRETFMSLVLFQ SNFGISFVSLSIWSKTCMRERFLDKLNDVSDKRYEYAEIORSRLRMIFVLP SL TLEATYRIWYISGPNQIPIIINILSHVVACTIQLSSWLTFRSLFIIVCLYKITCH LQTLRDDPACFASEITDVRSGALGEHQIKRNLRIVSHRFRFILLSLILVTQFM ALLTTTRASAVANIYEGELALCSLSVTGVICLRSAKITHKQSVTLAAKNVYC ATVDSFDLDEETPTGSIIESQVSLRGNAIETSDDEGEQDDLDNTHIPIYANTIS YQRQALVTYLENNKAGITVYGLVDRSWLNTFGIELALLLNLNKTIGLA"		
	10245..10394		
intron	/gene="AT4g22270"		
	/number=1		
exon	10395..10516		
	/gene="AT4g22270"		
	/number=2		
intron	10517..11326		
	/gene="AT4g22270"		
	/number=2		
trna	11168..11253		
	/gene="AT4g22270"		
	/note="Tyr-tRNA_(GTA)"		
exon	11327..11779		
	/gene="AT4g22270"		
	/number=3		
intron	11780..11867		
	/gene="AT4g22270"		
	/number=3		
exon	11868..12174		
	/gene="AT4g22270"		
	/number=4		
intron	12175..12253		
	/gene="AT4g22270"		



\* 82248 93854: contig of 11607 bp in length  
\* 93855 93954: gap of unknown length  
\* 93955 103236: contig of 9282 bp in length  
\* 103237 103336: gap of unknown length  
\* 103337 117502: contig of 14166 bp in length  
\* 117503 117602: gap of unknown length  
\* 117603 130162: contig of 12560 bp in length  
\* 130163 130262: gap of unknown length  
\* 130263 140708: contig of 10446 bp in length  
\* 140709 140808: gap of unknown length  
\* 140809 151652: contig of 10844 bp in length  
\* 151653 151753: gap of unknown length  
\* 151753 163824: contig of 12071 bp in length  
\* 163824 163923: gap of unknown length  
\* 163924 178730: contig of 14807 bp in length  
\* 178731 178830: gap of unknown length  
\* 178831 193909: contig of 15079 bp in length  
\* 193910 194009: gap of unknown length  
\* 194010 208684: contig of 14675 bp in length.

FEATURES

source  
1. .208684  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/chromosome="6"  
/clone\_lib="RPCI mouse BAC library 23"  
1. .2209  
/note="assembly\_fragment"  
2310. .4736  
/note="assembly\_fragment"  
4837. 7023  
/note="assembly\_fragment"  
7124. .9505  
/note="assembly\_fragment"  
9606. .14706  
/note="assembly\_fragment"  
14807. 18659  
/note="assembly\_fragment"  
18760. .25424  
/note="assembly\_fragment"  
25525. .33299  
/note="assembly\_fragment"  
33400. .39449  
/note="assembly\_fragment"  
39550. 48515  
/note="assembly\_fragment"  
48616. .55023  
/note="assembly\_fragment"  
55124. .64442  
/note="assembly\_fragment"  
64543. 73027  
/note="assembly\_fragment"  
73128. .82147  
/note="assembly\_fragment"  
82248. .93854  
/note="assembly\_fragment"  
93955. 103236  
/note="assembly\_fragment"  
103337. 117502  
/note="assembly\_fragment"  
117603. 130162  
/note="assembly\_fragment"  
130263. 140708  
/note="assembly\_fragment"  
140809. 151652  
/note="assembly\_fragment"  
151753. 163823  
/note="assembly\_fragment"  
163924. 178730  
/note="assembly\_fragment"  
clone\_end.T7  
vector\_side:right"

misc\_feature 178831. .193909  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:right"  
misc\_feature 194010. .208684  
/note="assembly\_fragment"  
BASE COUNT 54390 a 48824 c 48605 g 54542 t 2323 others  
ORIGIN  
Query Match 81.7%; Score 18.8; DB 2; Length 208684;  
Best Local Similarity 90.9%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 aaatccatgtaaagcagcagg 23  
||||| ||||| ||||| |||||  
Db 135605 AAATCCCTGTAAACCACGAGG 135584  
RESULT 9  
AC046148  
LOCUS  
DEFINITION  
AC046148 168524 bp DNA HTG  
Mus musculus chromosome 14 clone RP23-265D21, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 45 unordered pieces.  
AC046148  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 168524)  
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,  
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,  
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,  
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,  
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,  
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,  
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,  
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,  
Kovar,C., Liu,J., Liu,W., Louissegh,H., Lozado,R.J., Martin,R.,  
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,  
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,  
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,  
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgeson,A.,  
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,  
Worley,K. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 168524)  
Worley,K.C.  
Direct Submission  
Submitted (13-APR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 4, 2000 this sequence version replaced gi:9929612.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: MAEB  
Center clone name: RP23-265D21  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-primer Bodipy; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 132230 bases at least Q40  
Consensus quality: 149333 bases at least Q30  
Consensus quality: 156096 bases at least Q20  
Estimated insert size: 153798; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
\* 10854 10953: contig of 10853 bp in length  
\* 10954 20346: contig of 9393 bp in length  
\* 20347 20446: gap of unknown length  
\* 20447 30391: contig of 9945 bp in length  
\* 30392 30491: gap of unknown length  
\* 30492 39449: contig of 8858 bp in length  
\* 39450 39449: gap of unknown length  
\* 39450 47183: contig of 7734 bp in length  
\* 47184 47283: gap of unknown length  
\* 47284 53762: contig of 6479 bp in length  
\* 53763 53862: gap of unknown length  
\* 53863 58669: contig of 4807 bp in length  
\* 58670 58769: gap of unknown length  
\* 58770 65058: contig of 6289 bp in length  
\* 65059 65158: gap of unknown length  
\* 65159 69830: contig of 4672 bp in length  
\* 69831 69930: gap of unknown length  
\* 69931 73910: contig of 3980 bp in length  
\* 73911 74010: gap of unknown length  
\* 74011 78684: contig of 4674 bp in length  
\* 78685 78784: gap of unknown length  
\* 78785 83343: contig of 4559 bp in length  
\* 83344 83443: gap of unknown length  
\* 83444 87849: contig of 4406 bp in length  
\* 87850 87949: gap of unknown length  
\* 87950 92076: contig of 4127 bp in length  
\* 92077 92176: gap of unknown length  
\* 92177 97227: contig of 5051 bp in length  
\* 97228 97327: gap of unknown length  
\* 97328 101356: contig of 4029 bp in length  
\* 101357 101456: gap of unknown length  
\* 101457 104693: contig of 3237 bp in length  
\* 104694 104793: gap of unknown length  
\* 104794 108132: contig of 3339 bp in length  
\* 108133 108233: gap of unknown length  
\* 108233 111747: contig of 3515 bp in length  
\* 111748 111847: gap of unknown length  
\* 111848 115180: contig of 3333 bp in length  
\* 115181 115280: gap of unknown length  
\* 115281 118248: contig of 3288 bp in length  
\* 118249 118648: gap of unknown length  
\* 118649 121539: contig of 2891 bp in length  
\* 121540 121639: gap of unknown length  
\* 121640 124526: contig of 2887 bp in length  
\* 124527 124627: gap of unknown length  
\* 124627 127513: contig of 2886 bp in length  
\* 127513 127612: gap of unknown length  
\* 127613 131027: contig of 3415 bp in length  
\* 131028 131127: gap of unknown length  
\* 131128 133604: contig of 2477 bp in length  
\* 133605 133704: gap of unknown length  
\* 133705 135894: contig of 2190 bp in length  
\* 135895 135994: gap of unknown length  
\* 135995 137407: contig of 1413 bp in length  
\* 137408 137507: gap of unknown length  
\* 137508 139522: contig of 2015 bp in length  
\* 139523 139622: gap of unknown length  
\* 139623 141182: contig of 1560 bp in length  
\* 141183 141282: gap of unknown length  
\* 141283 143104: contig of 1822 bp in length

\* 143105 143204: gap of unknown length  
\* 143205 145597: contig of 2393 bp in length  
\* 145598 145697: gap of unknown length  
\* 145698 147940: contig of 2143 bp in length  
\* 147941 147940: gap of unknown length  
\* 150388 150388: contig of 2448 bp in length  
\* 150389 150488: gap of unknown length  
\* 150489 151853: contig of 1365 bp in length  
\* 151854 151953: gap of unknown length  
\* 151954 154894: contig of 2941 bp in length  
\* 154895 154994: gap of unknown length  
\* 154995 156000: contig of 1606 bp in length  
\* 156001 156700: gap of unknown length  
\* 156701 158135: contig of 1435 bp in length  
\* 158136 158235: gap of unknown length  
\* 158236 159267: contig of 1032 bp in length  
\* 159268 159367: gap of unknown length  
\* 159368 161428: contig of 2061 bp in length  
\* 161429 161528: gap of unknown length  
\* 161529 162993: contig of 1465 bp in length  
\* 162994 163093: gap of unknown length  
\* 163094 164386: contig of 1293 bp in length  
\* 164387 164486: gap of unknown length  
\* 164487 165850: contig of 1364 bp in length  
\* 165851 165950: gap of unknown length  
\* 165951 167364: contig of 1414 bp in length  
\* 167365 167464: gap of unknown length  
\* 167465 168524: contig of 1060 bp in length.

#### FEATURES

Location/Qualifiers  
1. .168524  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="14"  
/clone="RP23-245D21"

BASE COUNT 45461 a 36455 c 36516 g 45685 t 4407 others  
ORIGIN

Query Match 80.0%; Score 18.4; DB 2; Length 168524;  
Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaatccatgtataagcagcag 21  
|||||

Db 112558 AAATCCATGTATACACGAG 112577

RESULT 10  
AC040920/c

LOCUS

AC040920 172650 bp DNA HTG 11-APR-2000  
Homo sapiens chromosome 5 clone RP11-509M23 map 5, WORKING DRAFT  
SEQUENCE, 19 unordered pieces.

DEFINITION

AC040920.1 GI:7534094

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 172650)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 5, clone RP11-509M23  
Unpublished  
2 (bases 1 to 172650)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
 Levine, R., Liu, G., Locke, C., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrim, J., Menus, L., Minova, T., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Ollivar, F.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody M.  
 Direct Submission

# TITLE JOURNAL

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8553

Center clone name: 509\_M\_23

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 163069 bases at least Q40

Consensus quality: 167976 bases at least Q30

Consensus quality: 169595 bases at least Q20

Insert size: 170000; agarose-fp

Quality coverage: 4.7 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 19 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 422: contig of 422 bp in length  
 423 522: gap of 100 bp  
 523 2269: contig of 1747 bp in length  
 2270 2369: gap of 100 bp  
 2370 4946: contig of 2577 bp in length  
 4947 5046: gap of 100 bp  
 5047 7162: contig of 2116 bp in length  
 7163 7262: gap of 100 bp  
 7263 9787: contig of 2525 bp in length  
 9788 9887: gap of 100 bp  
 9888 13694: contig of 3807 bp in length  
 13695 13794: gap of 100 bp  
 13795 18462: contig of 4688 bp in length  
 18483 18582: gap of 100 bp  
 18583 24380: contig of 5798 bp in length  
 24381 24480: gap of 100 bp  
 24481 30597: contig of 6117 bp in length  
 30598 30697: gap of 100 bp  
 30698 36241: contig of 5544 bp in length  
 36242 36341: gap of 100 bp  
 36342 43886: contig of 7545 bp in length  
 43887 43986: gap of 100 bp  
 43987 53642: contig of 9656 bp in length  
 53643 53742: gap of 100 bp  
 53743 65843: contig of 12101 bp in length  
 65844 65944: gap of 100 bp  
 65944 82865: contig of 16922 bp in length

\* 82866 82965: gap of 100 bp  
 82966 97230: contig of 14265 bp in length  
 97231 97330: gap of 100 bp  
 97331 114682: contig of 17352 bp in length  
 114683 114782: gap of 100 bp  
 114783 132889: contig of 17507 bp in length  
 132890 132889: gap of 100 bp  
 13290 151664: contig of 19275 bp in length  
 151665 151764: gap of 100 bp  
 151765 172650: contig of 20886 bp in length.  
 172650 Location/Qualifiers  
 1. .172650

## FEATURES

source

misc\_feature

1. .422

/note="assembly\_fragment"

clone\_end:T7

vector\_side:left

523. .2269

/note="assembly\_fragment"

2370. .4946

/note="assembly\_fragment"

5047. .7162

/note="assembly\_fragment"

7263. .9787

/note="assembly\_fragment"

9888. .13694

/note="assembly\_fragment"

13795. .18482

/note="assembly\_fragment"

18583. .24380

/note="assembly\_fragment"

24481. .30597

/note="assembly\_fragment"

30698. .36241

/note="assembly\_fragment"

36342. .43886

/note="assembly\_fragment"

43987. .53642

/note="assembly\_fragment"

53743. .65843

/note="assembly\_fragment"

65944. .82865

/note="assembly\_fragment"

82966. .97230

/note="assembly\_fragment"

97331. .114682

/note="assembly\_fragment"

114783. .132289

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left

132390. .151664

/note="assembly\_fragment"

151765. .172650

/note="assembly\_fragment"

BASE COUNT 57057 a 28688 c 29466 g 55635 t 1804 others

ORIGIN

Query Match 80.0%; Score 18.4; DB 2; Length 172650;  
 Best Local Similarity 95.0%; Pred. No. 1.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 aaatccatgtaaagcagcag 21

|||||  
 Db 142729 AAAGCATGTAAAGCAGCAG 142710



```

RESULT 11
AF118429          535 bp      DNA          VRT          10-JAN-2001
LOCUS             AF118429          microsatellite NVH Fp86-2 sequence.
DEFINITION        Falco peregrinus microsatellite NVH Fp86-2 sequence.
ACCESSION         AF118429
VERSION           AF118429.1  GI:5840932
KEYWORDS
SOURCE
ORGANISM          Falco peregrinus.
REFERENCE
AUTHORS           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE             Archosauria; Aves; Neognathae; Falconiformes; Falconidae; Falco.
JOURNAL           1 (bases 1 to 535)
MEDLINE           Nesje,M., Roed,K.H., Lifjeld,J.T., Lindberg,P. and Steen,O.F.
PUBMED            Genetic relationships in the peregrine falcon (Falco peregrinus)
AUTHORS           analysed by microsatellite DNA markers
TITLE             Mol. Ecol. 9 (1), 53-60 (2000)
JOURNAL           20117586
MEDLINE           10652075
PUBMED            Nesje,M. and Roed,K.H.
AUTHORS           Direct Submission
TITLE             Submitted (07-JAN-1999) MGA, Genetics, Norwegian College of
JOURNAL           Veterinary Medicine, P.O. Box 8146 Dep., Oslo 0033, Norway
FEATURES
source            Location/Qualifiers
                1..535
                /organism="Falco peregrinus"
                /db_xref="taxon:9954"
repeat_region    1..535
                /note="microsatellite NVH Fp86-2"
                /rpt_type=tandem
BASE COUNT        168 a 110 c 117 g 140 t
ORIGIN
Query Match      79.1%; Score 18.2; DB 5; Length 535;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 gaaatccatgtaaagcagcaggg 23
| ||||| ||||| ||||| |||||
Db 365 GGAATCCATGACACGACGAGG 407

RESULT 12
AC005083/c
LOCUS             AC005083          146285 bp      DNA          PRI          21-DEC-1999
DEFINITION        Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.
ACCESSION         AC005083
VERSION           AC005083.1  GI:4150930
KEYWORDS          HTG.
SOURCE            human.
ORGANISM          Homo sapiens
REFERENCE
AUTHORS           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL           Madsen,C. and Blair,T.
AUTHORS           The sequence of Homo sapiens BAC clone CTA-281G5
TITLE             Unpublished
JOURNAL           2 (bases 1 to 146285)
MEDLINE           Waterston,R.H.
PUBMED            Direct Submission
AUTHORS           University (12-JUN-1998) Genome Sequencing Center, Washington
TITLE             Submitted (12-JUN-1998) Genome Sequencing Center, Washington
JOURNAL           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS           3 (bases 1 to 146285)
TITLE             Waterston,R.
JOURNAL           Direct Submission
AUTHORS           Submitted (12-JAN-1999) Department of Genetics, Washington
TITLE             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS           4 (bases 1 to 146285)
TITLE             Waterston,R.
JOURNAL           Direct Submission
AUTHORS           Submitted (21-DEC-1999) Department of Genetics, Washington

```

## COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jan 12, 1999 this sequence version replaced gi:3212908.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: H\_RG281G05  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send [mailto:egreen@nhgri.nih.gov](mailto:mailto:egreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

Clone CTA-281G5 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
 VECTOR: pBelOBAC11  
 Selection: chloramphenicol

## NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-281G5; actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an E. coli transposon from 10982 to 12345 that is not represented in the submitted sequence.

## FEATURES

## source

```

1..146285
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p15-p21"
/clone="CTA-281G5"
/clone_lib="CITB-HS-A"
322..358
/rpt_family="(TAAAA)n"
808..952
/rpt_family="MER1_type"
1172..1194
/rpt_family="AT_rich"
1264..1720
/rpt_family="L1"
1848..1881
/rpt_family="AT_rich"
1882..2171
/rpt_family="Alu"
3782..3804
/rpt_family="AT_rich"
4794..4826
/rpt_family="AT_rich"
5764..6034

```



```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAAW
Center clone name: RP23-361K18
----- Summary Statistics
Sequencing vector: M13: L08821
Chemistry: Dye-terminator Big Dye: 70% of reads
Chemistry: Dye-terminator Big Dye: 30% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 137454 bases at least Q40
Consensus quality: 161084 bases at least Q30
Consensus quality: 170194 bases at least Q20
Estimated insert size: 164611; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 10583: contig of 10583 bp in length
* 10584: gap of unknown length
* 10684: contig of 9744 bp in length
* 20428: gap of unknown length
* 20527: gap of unknown length
* 29063: contig of 8536 bp in length
* 29084: gap of unknown length
* 29164: contig of 7048 bp in length
* 36211: gap of unknown length
* 36312: contig of 5943 bp in length
* 36312: contig of 5943 bp in length
* 42255: gap of unknown length
* 42355: contig of 5338 bp in length
* 47692: contig of 5338 bp in length
* 47693: gap of unknown length
* 51859: contig of 4067 bp in length
* 51860: gap of unknown length
* 51950: gap of unknown length
* 55000: contig of 3041 bp in length
* 55001: gap of unknown length
* 55100: gap of unknown length
* 59766: contig of 4666 bp in length
* 59767: gap of unknown length
* 59866: gap of unknown length
* 59867: contig of 2791 bp in length
* 62658: gap of unknown length
* 62757: gap of unknown length
* 62758: contig of 4413 bp in length
* 67171: gap of unknown length
* 67270: gap of unknown length
* 67271: contig of 2929 bp in length
* 70199: gap of unknown length
* 70299: gap of unknown length
* 70300: contig of 3826 bp in length
* 74125: gap of unknown length
* 74225: gap of unknown length
* 77307: contig of 3082 bp in length
* 77407: gap of unknown length
* 79603: contig of 2196 bp in length
* 79703: gap of unknown length
* 79704: contig of 2982 bp in length
* 82685: gap of unknown length
* 82786: contig of 3599 bp in length
* 86385: gap of unknown length
* 86484: contig of 3583 bp in length
* 90067: gap of unknown length
* 90167: gap of unknown length
* 93018: contig of 2851 bp in length
* 93118: gap of unknown length
* 96993: contig of 3875 bp in length
* 97093: gap of unknown length
* 100012: contig of 2919 bp in length
* 100112: gap of unknown length
* 103857: contig of 3745 bp in length

```

```

* 103858: gap of unknown length
* 103958: contig of 3565 bp in length
* 107522: gap of unknown length
* 107523: gap of unknown length
* 107623: contig of 2424 bp in length
* 110046: gap of unknown length
* 110047: gap of unknown length
* 110147: contig of 2952 bp in length
* 113098: gap of unknown length
* 113198: gap of unknown length
* 113199: contig of 3372 bp in length
* 116570: gap of unknown length
* 116571: gap of unknown length
* 116670: gap of unknown length
* 116671: contig of 2343 bp in length
* 119013: gap of unknown length
* 119014: gap of unknown length
* 119114: contig of 2940 bp in length
* 122053: gap of unknown length
* 122153: gap of unknown length
* 122154: contig of 2806 bp in length
* 124960: gap of unknown length
* 125059: gap of unknown length
* 125060: contig of 2109 bp in length
* 127168: gap of unknown length
* 127268: gap of unknown length
* 127269: contig of 2336 bp in length
* 129605: gap of unknown length
* 129704: gap of unknown length
* 129705: contig of 3451 bp in length
* 13155: gap of unknown length
* 13255: gap of unknown length
* 133156: contig of 3353 bp in length
* 133255: gap of unknown length
* 133256: gap of unknown length
* 136609: gap of unknown length
* 136709: contig of 1419 bp in length
* 138127: gap of unknown length
* 138128: contig of 2776 bp in length
* 138228: gap of unknown length
* 141003: contig of 2776 bp in length
* 141104: gap of unknown length
* 141104: contig of 2591 bp in length
* 143694: gap of unknown length
* 143695: gap of unknown length
* 143795: contig of 1117 bp in length
* 144911: gap of unknown length
* 145011: gap of unknown length
* 145012: contig of 1349 bp in length
* 146360: gap of unknown length
* 146361: gap of unknown length
* 146460: gap of unknown length
* 146461: contig of 2530 bp in length
* 148990: gap of unknown length
* 148991: contig of 1302 bp in length
* 149091: gap of unknown length
* 150392: contig of 1710 bp in length
* 150492: gap of unknown length
* 150991: contig of 1710 bp in length
* 152202: gap of unknown length
* 152302: gap of unknown length
* 152303: contig of 2092 bp in length
* 154394: gap of unknown length
* 154395: contig of 1921 bp in length
* 154949: gap of unknown length
* 156115: gap of unknown length
* 156116: contig of 1458 bp in length
* 157973: gap of unknown length
* 157974: contig of 1471 bp in length
* 158073: gap of unknown length
* 159544: gap of unknown length
* 159545: contig of 1467 bp in length
* 161111: gap of unknown length
* 161112: contig of 1612 bp in length
* 162823: gap of unknown length
* 162923: gap of unknown length
* 164564: contig of 1641 bp in length
* 164664: gap of unknown length
* 164665: contig of 1431 bp in length
* 166095: gap of unknown length
* 166195: gap of unknown length
* 166196: contig of 1776 bp in length
* 167971: gap of unknown length
* 167972: contig of 1235 bp in length
* 168071: gap of unknown length
* 168072: contig of 1288 bp in length
* 169406: gap of unknown length
* 169407: contig of 1288 bp in length
* 170694: gap of unknown length
* 170695: contig of 1559 bp in length
* 172353: gap of unknown length
* 172354: contig of 1010 bp in length
* 173463: gap of unknown length
* 173464: contig of 1339 bp in length
* 174902: gap of unknown length
* 174903: contig of 1924 bp in length
* 175002: gap of unknown length
* 175003: contig of 1924 bp in length
* 176926: gap of unknown length
* 176927: contig of 1483 bp in length
* 177027: gap of unknown length
* 178509: gap of unknown length
* 178610: contig of 1225 bp in length
* 179834: gap of unknown length
* 179835: contig of 1225 bp in length

```

\* 179935 181235: contig of 1301 bp in length  
 \* 181236 181335: gap of unknown length  
 \* 181336 182487: contig of 1152 bp in length  
 \* 182488 182587: gap of unknown length  
 \* 182588 183634: contig of 1047 bp in length.

## FEATURES

source

1. 183634  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP23-361K18"

BASE COUNT 48869 a 39982 c 39763 g 49962 t 6058 others  
 ORIGIN

Query Match 79.1%; Score 18.2; DB 2; Length 183634;  
 Best Local Similarity 87.0%; Pred No. 2.1e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaatccatgttaagcagcaggg 23  
 ||||| || ||||| |||||

Db 60779 GAATCCCTGCTAGCAGCAGG 60757

## RESULT 14

AC021443/c

LOCUS

DEFINITION

Homo sapiens chromosome 11 clone RP11-692M12 map 11, WORKING DRAFT  
 SEQUENCE, 25 unordered pieces.

ACCESSION

AC021443

VERSION

AC021443.8

KEYWORDS

HTGS; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 228809)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,

Boguslavsky, L., Bouckhvalter, B., Brown, A., Burkett, G., Castle, A.,

Choquel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,

DeArelano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,

Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,

Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lander, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G., Locke, K.,

Macdonald, P., Marquis, J., McEwan, P., McGurk, A., McKernan, K.,

McPheeters, R., Meldrum, J., Menes, L., Morrow, J., Naylor, J.,

Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K.,

Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,

Roy, A., Santos, R., Severi, P., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,

Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,

Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 31, 2001 this sequence version replaced gi:11560245.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

Sequencing vector: M13; M77815; 43% of reads  
 Sequencing vector: Plasmid; L08752; 1% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 214328 bases at least Q40  
 Consensus quality: 219470 bases at least Q30  
 Consensus quality: 222925 bases at least Q20  
 Insert size: 198000; agarose-fp  
 Insert size: 226409; sum-of-contigs  
 Quality coverage: 10.4 in Q20.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 25 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 3785: contig of 3785 bp in length  
 3786 3885: gap of 100 bp  
 3886 5107: contig of 1222 bp in length  
 5108 5207: gap of 100 bp  
 5208 6375: contig of 1168 bp in length  
 6376 6475: gap of 100 bp  
 6476 7525: contig of 1050 bp in length  
 7526 7625: gap of 100 bp  
 7626 8701: contig of 1076 bp in length  
 8702 8801: gap of 100 bp  
 8802 10296: contig of 1495 bp in length  
 10297 10396: gap of 100 bp  
 10397 11470: contig of 1074 bp in length  
 11471 11570: gap of 100 bp  
 11571 12684: contig of 1114 bp in length  
 12685 12784: gap of 100 bp  
 12785 14999: contig of 2215 bp in length  
 15000 15099: gap of 100 bp  
 15100 18672: contig of 3573 bp in length  
 18673 18772: gap of 100 bp  
 18773 22303: contig of 3531 bp in length  
 22304 22403: gap of 100 bp  
 22404 28525: contig of 6122 bp in length  
 28526 28625: gap of 100 bp  
 28626 36093: contig of 7468 bp in length  
 36094 36193: gap of 100 bp  
 36194 45198: contig of 9005 bp in length  
 45199 45298: gap of 100 bp  
 45299 56916: contig of 11618 bp in length  
 56917 57016: gap of 100 bp  
 57017 65985: contig of 8969 bp in length  
 65986 66085: gap of 100 bp  
 66086 78887: contig of 12802 bp in length  
 78888 78987: gap of 100 bp  
 78988 94277: contig of 15290 bp in length  
 94278 94377: gap of 100 bp  
 94378 109994: contig of 15617 bp in length  
 109995 110094: gap of 100 bp  
 110095 121149: contig of 11055 bp in length  
 121150 121249: gap of 100 bp  
 121250 138125: contig of 16876 bp in length  
 138126 138225: gap of 100 bp  
 138226 155392: contig of 17167 bp in length  
 155393 155492: gap of 100 bp  
 155493 179217: contig of 23725 bp in length  
 179218 179317: gap of 100 bp  
 179318 202705: contig of 23388 bp in length  
 202706 202805: gap of 100 bp  
 202806 228809: contig of 26004 bp in length.

## FEATURES

source

1. 228809  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:38 ; Search time 716.55 Seconds  
(without alignments)  
27.519 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23

Sequence: 1 gaaatccatgtaaacgcagcagg 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	AAD07000	PCR primer B01, to
2	23	100.0	415	AAD06997	Right (5') border
3	23	100.0	416	AAD06999	Left (3') border f
4	18.8	81.7	1152	AAC46330	Arabidopsis thalia
5	17.4	75.7	1567	AAAC79889	Human secreted pro
6	17.2	74.8	1153	AAC39525	Arabidopsis thalia
7	17.2	74.8	2857	AAH18005	Human cDNA sequenc
8	17.2	74.8	4495	AAH18005	Human polynucleoti
9	17.2	74.8	4775	AAI60253	Human polynucleoti
10	17.2	74.8	8826	AAT74989	Saccharomyces cere
11	17.2	74.8	9621	AAT94548	Mutant YLR087c gen

c 12	16.8	73.0	602	22	AAH29132	Drosophila melanog
c 13	16.8	73.0	1752	22	AAF57408	Human p95 protein
c 14	16.8	73.0	4388	18	AAT60073	Selective marker g
c 15	16.6	72.2	281	14	AAQ60834	Human brain Expres
c 16	16.6	72.2	292	16	AAT25120	Human gene signatu
c 17	16.6	72.2	292	20	AAV88563	EST clone EY281.
c 18	16.6	72.2	292	21	AAZ46937	Human saccharide-t
c 19	16.6	72.2	549	19	AAV49585	Human stomach canc
c 20	16.6	72.2	552	21	AAQ39940	Human TANGO 183 co
c 21	16.6	72.2	575	22	AAF93941	Primer specific fo
c 22	16.6	72.2	692	22	AAH32272	Human olfactory re
c 23	16.6	72.2	917	21	AAC76546	Human ORFX ORF2101
c 24	16.6	72.2	1026	21	AAC51028	Arabidopsis thalia
c 25	16.6	72.2	1056	21	AAF51420	Arabidopsis thalia
c 26	16.6	72.2	1499	22	AAF93749	Human cDNA encodin
c 27	16.6	72.2	1499	22	AAF29357	Anyloid-beta prote
c 28	16.6	72.2	1504	19	AAV49584	Human stomach canc
c 29	16.6	72.2	1563	22	AAH64828	Human secreted pro
c 30	16.6	72.2	1564	21	AAH37041	Human PRO1248 (UNQ
c 31	16.6	72.2	1564	22	AAF54249	DNA encoding prote
c 32	16.6	72.2	1565	21	AAQ39939	Human TANGO 183 cD
c 33	16.6	72.2	1571	20	AAZ41983	Human endometrium
c 34	16.6	72.2	1645	21	AAF21831	Human breast and o
c 35	16.6	72.2	1821	21	AAC51418	Arabidopsis thalia
c 36	16.6	72.2	1851	21	AAC51029	Arabidopsis thalia
c 37	16.6	72.2	1929	22	AAC85075	Atherosclerosis-as
c 38	16.6	72.2	1962	22	AAF33128	Human secreted pro
c 39	16.6	72.2	2370	22	AAF26956	Human cancer assoc
c 40	16.6	72.2	2380	21	AAF76858	Human ORFX ORF2413
c 41	16.6	72.2	2383	22	AAH22608	Human cDNA encodin
c 42	16.6	72.2	2393	22	AAH98580	Human EST-derived
c 43	16.6	72.2	2393	22	AAH22844	Human cDNA encodin
c 44	16.6	72.2	2402	21	AAZ33327	Human secreted pro
c 45	16.6	72.2	2460	20	AAH58671	Human organic cati

#### ALIGNMENTS

RESULT 1

AAD07000

ID AAD07000 standard; DNA; 23 BP.

XX

AC AAD07000;

XX

DT 06-AUG-2001 (first entry)

XX

DE PCR primer B01, to recognise foreign DNA and flanking sequence of MS-B2.

XX

KW MS-B2 elite event; transgenic Brassica plant; transformation event;

XX male-sterility gene; PCR primer; ss.

XX

OS Unidentified.

XX

PN WO200131042-A2.

XX

PD 03-MAY-2001.

XX

PF 26-OCT-2000; 2000WO-EP10680.

XX

PR 29-OCT-1999; 99US-0430497.

XX

PA (AVET ) AVENTIS CROPS SCIENCE NV.

 XX |  |  |  |  |  || PI | Weston B, | De Beuckeleer M; |  |  |  |
XX					
DR	WPI;	2001-300517/31.			
XX					
PT	Transgenic Brassica plants, seeds, cells or tissues, characterized by				
PT	harboring specific transformation events, particularly by presence of				
XX	male-sterility gene, at specific location in its genome -				
PS	Claim 1; Page 33; 53pp; English.				

```

XX CC The present invention relates to a transgenic Brassica plant or its
XX CC seed, cells or tissues, characterised by harbouring a specific
XX CC transformation event, particularly by the presence of a male-sterility
XX CC gene, at a specific location in the Brassica genome. Transgenic
XX CC Brassica plant is useful for producing a hybrid seed by crossing the
XX CC transgenic plant with a male-fertile Brassica plant and harvesting the
XX CC hybrid seed from the transgenic Brassica plant.
XX CC The present sequence is PCR primer which is used to recognise foreign
XX CC DNA and a flanking sequence of elite event MS-B2.
XX SQ Sequence 23 BP; 9 A; 4 C; 7 G; 3 T; 0 other;

Query Match      100.0%; Score 23; DB 22; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23
Db 1 gaaatccatgtaaagcagcagg 23

RESULT 2
AAD06997
ID AAD06997 standard; DNA; 415 BP.
XX AC AAD06997;
XX DT 06-AUG-2001 (first entry)
XX DE Right (5') border flanking region of elite event MS-B2.
XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;
XX KW male-sterility gene; ds.
XX OS Chimeric - Agrobacterium sp.
XX OS Chimeric - Brassica sp.
XX FH Key Location/Qualifiers
XX FT misc_feature 1..234
XX FT /*tag= a
XX FT /note= "Corresponds to plant DNA"
XX FT misc_feature 235..415
XX FT /*tag= b
XX FT /note= "Corresponds to T-DNA"
XX PN WO200131042-A2.
XX XX 03-MAY-2001.
XX PD 26-OCT-2000; 2000WO-EP10680.
XX PF 29-OCT-1999; 99US-0430497.
XX PR (AVET ) AVENTIS CROPS SCIENCE NV.
XX PA Weston B, De Beuckeleer M;
XX PI WPI; 2001-300517/31.
XX DR Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX PT harboring specific transformation events, particularly by presence of
XX PT male-sterility gene, at specific location in its genome -
XX PS Claim 11; Page 51; 53pp; English.
XX CC The present invention relates to a transgenic Brassica plant or its
XX CC seed, cells or tissues, characterised by harbouring a specific
XX CC transformation event, particularly by the presence of a male-sterility
XX CC gene, at a specific location in the Brassica genome. Transgenic
XX CC Brassica plant is useful for producing a hybrid seed by crossing the
XX CC transgenic plant with a male-fertile Brassica plant and harvesting the

```

```

CC hybrid seed from the transgenic Brassica plant.
CC The present sequence is right (5') border flanking region of elite event
CC MS-B2.
XX SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match      100.0%; Score 23; DB 22; Length 415;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23
Db 152 gaaatccatgtaaagcagcagg 174

RESULT 3
AAD06999/c
ID AAD06999 standard; DNA; 416 BP.
XX AC AAD06999;
XX DT 06-AUG-2001 (first entry)
XX DE Left (3') border flanking region of elite event MS-B2.
XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;
XX KW male-sterility gene; ds.
XX OS Chimeric - Agrobacterium sp.
XX OS Chimeric - Brassica sp.
XX FH Key Location/Qualifiers
XX FT misc_feature 1..193
XX FT /*tag= a
XX FT /note= "Corresponds to T-DNA"
XX FT misc_feature 194..416
XX FT /*tag= b
XX FT /note= "Corresponds to plant DNA"
XX PN WO200131042-A2.
XX XX 03-MAY-2001.
XX PD 26-OCT-2000; 2000WO-EP10680.
XX PF 29-OCT-1999; 99US-0430497.
XX PR (AVET ) AVENTIS CROPS SCIENCE NV.
XX PA Weston B, De Beuckeleer M;
XX PI WPI; 2001-300517/31.
XX DR Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX PT harboring specific transformation events, particularly by presence of
XX PT male-sterility gene, at specific location in its genome -
XX PS Claim 11; Page 52; 53pp; English.
XX CC The present invention relates to a transgenic Brassica plant or its
XX CC seed, cells or tissues, characterised by harbouring a specific
XX CC transformation event, particularly by the presence of a male-sterility
XX CC gene, at a specific location in the Brassica genome. Transgenic
XX CC Brassica plant is useful for producing a hybrid seed by crossing the
XX CC transgenic plant with a male-fertile Brassica plant and harvesting the
XX CC hybrid seed from the transgenic Brassica plant.
XX CC The present sequence is left (3') border flanking region of elite event
XX CC MS-B2.
XX SQ Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;

```

Query Match	100.0%;	Score 23;	DB 22;	Length 416;
Best Local Similarity	100.0%;	Pred. No. 0.14;		
Matches 23;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	gaatccatgtataagcagcagg 23		
Db	263	GAATCCATGTAAAGCAGCAGG 241		
RESULT	4			
AAC46330				
ID	AAC46330	standard; DNA; 1152 BP.		
XX	AC	AAC46330;		
XX	DT	18-OCT-2000 (first entry)		
XX	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 49743.		
XX	KW	Hybridisation assay; genetic mapping; gene expression control;		
KW	KW	protein identification; signal transduction pathway;		
KW	KW	metabolic pathway; promoter; termination sequence; ss.		
XX	OS	Arabidopsis thaliana.		
XX	PN	EP1033405-A2.		
XX	PD	06-SEP-2000.		
PF	25-FEB-2000;	2000EP-0301439.		
XX	25-FEB-1999;	99US-0121825.		
PR	05-MAR-1999;	99US-0123180.		
PR	09-MAR-1999;	99US-0123548.		
PR	23-MAR-1999;	99US-0125788.		
PR	25-MAR-1999;	99US-0126284.		
PR	29-MAR-1999;	99US-0126785.		
PR	01-APR-1999;	99US-0127462.		
PR	06-APR-1999;	99US-0128234.		
PR	08-APR-1999;	99US-0128714.		
PR	16-APR-1999;	99US-0129845.		
PR	19-APR-1999;	99US-0130077.		
PR	21-APR-1999;	99US-0130449.		
PR	23-APR-1999;	99US-0130510.		
PR	28-APR-1999;	99US-0130891.		
PR	28-APR-1999;	99US-0131449.		
PR	30-APR-1999;	99US-0132048.		
PR	30-APR-1999;	99US-0132407.		
PR	04-MAY-1999;	99US-0132484.		
PR	05-MAY-1999;	99US-0132485.		
PR	06-MAY-1999;	99US-0132486.		
PR	06-MAY-1999;	99US-0132487.		
PR	07-MAY-1999;	99US-0132863.		
PR	11-MAY-1999;	99US-0134236.		
PR	14-MAY-1999;	99US-0134218.		
PR	14-MAY-1999;	99US-0134219.		
PR	14-MAY-1999;	99US-0134221.		
PR	14-MAY-1999;	99US-0134370.		
PR	18-MAY-1999;	99US-0134768.		
PR	19-MAY-1999;	99US-0134941.		
PR	20-MAY-1999;	99US-0135124.		
PR	21-MAY-1999;	99US-0135353.		
PR	24-MAY-1999;	99US-0135629.		
PR	25-MAY-1999;	99US-0136021.		
PR	27-MAY-1999;	99US-0136392.		
PR	28-MAY-1999;	99US-0136782.		
PR	01-JUN-1999;	99US-0137222.		
PR	03-JUN-1999;	99US-0137528.		
PR	04-JUN-1999;	99US-0137502.		
PR	07-JUN-1999;	99US-0137724.		
PR	08-JUN-1999;	99US-0138094.		
PR	10-JUN-1999;	99US-0138540.		
PR	10-JUN-1999;	99US-0138847.		
PR	14-JUN-1999;	99US-0139119.		
PR	16-JUN-1999;	99US-0139452.		
PR	16-JUN-1999;	99US-0139453.		
PR	17-JUN-1999;	99US-0139492.		
PR	18-JUN-1999;	99US-0139454.		
PR	18-JUN-1999;	99US-0139455.		
PR	18-JUN-1999;	99US-0139456.		
PR	18-JUN-1999;	99US-0139457.		
PR	18-JUN-1999;	99US-0139458.		
PR	18-JUN-1999;	99US-0139459.		
PR	18-JUN-1999;	99US-0139460.		
PR	18-JUN-1999;	99US-0139461.		
PR	18-JUN-1999;	99US-0139462.		
PR	18-JUN-1999;	99US-0139463.		
PR	18-JUN-1999;	99US-0139750.		
PR	18-JUN-1999;	99US-0139763.		
PR	21-JUN-1999;	99US-0139817.		
PR	22-JUN-1999;	99US-0139899.		
PR	23-JUN-1999;	99US-0140353.		
PR	23-JUN-1999;	99US-0140354.		
PR	24-JUN-1999;	99US-0140695.		
PR	28-JUN-1999;	99US-0140823.		
PR	29-JUN-1999;	99US-0140991.		
PR	30-JUN-1999;	99US-0141287.		
PR	01-JUL-1999;	99US-0141842.		
PR	01-JUL-1999;	99US-0142154.		
PR	02-JUL-1999;	99US-0142055.		
PR	06-JUL-1999;	99US-0142390.		
PR	08-JUL-1999;	99US-0142803.		
PR	09-JUL-1999;	99US-0142920.		
PR	12-JUL-1999;	99US-0142977.		
PR	13-JUL-1999;	99US-0143542.		
PR	14-JUL-1999;	99US-0143624.		
PR	15-JUL-1999;	99US-0144005.		
PR	16-JUL-1999;	99US-0144085.		
PR	16-JUL-1999;	99US-0144086.		
PR	19-JUL-1999;	99US-0144331.		
PR	19-JUL-1999;	99US-0144332.		
PR	19-JUL-1999;	99US-0144333.		
PR	19-JUL-1999;	99US-0144334.		
PR	19-JUL-1999;	99US-0144335.		
PR	20-JUL-1999;	99US-0144352.		
PR	20-JUL-1999;	99US-0144632.		
PR	20-JUL-1999;	99US-0144884.		
PR	21-JUL-1999;	99US-0145086.		
PR	21-JUL-1999;	99US-0145088.		
PR	22-JUL-1999;	99US-0145087.		
PR	22-JUL-1999;	99US-0145089.		
PR	22-JUL-1999;	99US-0145192.		
PR	23-JUL-1999;	99US-0145145.		
PR	23-JUL-1999;	99US-0145218.		
PR	23-JUL-1999;	99US-0145224.		
PR	26-JUL-1999;	99US-0145276.		
PR	27-JUL-1999;	99US-0145913.		
PR	27-JUL-1999;	99US-0145918.		
PR	27-JUL-1999;	99US-0145919.		
PR	28-JUL-1999;	99US-0145951.		
PR	02-AUG-1999;	99US-0146386.		
PR	02-AUG-1999;	99US-0146388.		
PR	02-AUG-1999;	99US-0146389.		
PR	03-AUG-1999;	99US-0147038.		
PR	04-AUG-1999;	99US-0147204.		
PR	04-AUG-1999;	99US-0147302.		
PR	05-AUG-1999;	99US-0147192.		
PR	05-AUG-1999;	99US-0147260.		
PR	06-AUG-1999;	99US-0147303.		
PR	06-AUG-1999;	99US-0147416.		
PR	09-AUG-1999;	99US-0147493.		
PR	09-AUG-1999;	99US-0147935.		
PR	10-AUG-1999;	99US-0148171.		

```
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      81.7%; Score 18.8; DB 21; Length 1152;
Best Local Similarity 90.9%; Pred. NO. 15;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aaatccatgtaagcagcaggg 23
    |||||
Db 1088 aaatccatgtaagcagtaagg 1109
```

## RESULT 5

AAC79889

ID AAC79889 standard; cDNA; 1567 BP.

XX AC AAC79889;

XX DT 09-FEB-2001 (first entry)

XX DE Human secreted protein encoding cDNA for gene 41.

XX KW Human; secreted protein; cytostatic; antiarthritic; antiasthmatic;  
XX KW immunosuppressive; antiarteriosclerotic; antiinflammatory; nootropic;  
XX KW neuroprotective; antidiabetic; tranquiliser; vulnerary; antibacterial;  
XX KW antipsoriatic; antiarrhythmic; antirheumatic; cardiant; anti-HIV;  
XX KW autoimmune disorder; allergic condition; cardiovascular disorder;  
XX KW cancer; neurological disease; tissue repair; ss.

XX OS Homo sapiens.

XX PN WO200055176-A2.

XX PD 21-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06057.

XX PR 12-MAR-1999; 99US-0124142.

XX PR 11-JUN-1999; 99US-0138597.

XX PR 03-DEC-1999; 99US-0168666.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM, Komatsoulis G;

XX P-PSDB; AAB44870.

PT Novel 49 human secreted proteins useful for diagnosis, prevention and  
PT treatment of disorders including neurological, cell proliferative,  
PT cardiovascular, and autoimmune/inflammatory disorders and microbial  
PT infections -

PS Claim la; Page 356-357; 405pp; English.

XX This invention describes a novel isolated polypeptide (I) comprising an  
XX amino acid sequence at least 95 % identical to a polypeptide sequence  
XX selected from 49 polypeptides encoded by polynucleotide sequences  
XX included in American Type Culture Collection (ATCC) deposit number  
XX 203917, defined in the specification. The products of the invention have  
XX cytostatic, antiarthritic, antiasthmatic, immunosuppressive, nootropic,  
XX antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,  
XX tranquiliser, vulnerary, antibacterial, antipsoriatic, antiarrhythmic,  
XX antirheumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)  
XX encoding (I) is useful for preventing, treating or ameliorating a medical  
XX condition and for diagnosing a pathological condition or susceptibility  
XX to the condition. (I) is useful for identifying a binding partner which  
XX affects the activity of the polypeptide and for identifying an activity  
XX in a biological sample. (I), (II) or an antibody (IV) specific to (I) is  
XX also useful for treating or preventing a disease, disorder or condition  
XX associated with aberrant expression of (I). Diseases treated or diagnosed  
XX include immune disorders such as autoimmune diseases, blood protein  
XX disorders, anemia, allergic reactions and conditions such as asthma,  
XX organ rejection or graft-versus-host disease, inflammation, hyper  
XX proliferative disorders, cardiovascular disorders such as arterioarterial  
XX fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ  
XX regeneration, cancer, neovascular glaucoma, diabetic retinopathy,  
XX rheumatoid arthritis, psoriasis, diseases associated with increased  
XX apoptosis that include acquired immunodeficiency syndrome (AIDS),  
XX neurological diseases such as Parkinson's disease, viral, bacterial,  
XX fungal or parasitic diseases. They are also used to repair, replace or  
XX protect tissue damage by congenital defects, to treat trauma, in surgery,  
XX including cosmetic plastic surgery, to treat fibrosis, reperfusion injury



CC or systemic cytokine damage, to stimulate chondrocyte growth, to prevent  
CC skin aging due to sunburn, to change a mammal's mental state or physical  
CC state by influencing biorhythms, cardiac rhythms, depression, memory,  
CC stress and for accelerating wound healing. (I), (II) and/or their agonist  
CC or antagonist are useful as food additives or preservatives to increase  
CC or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamin, mineral or other nutritional components. (I) is  
CC useful for screening therapeutic compounds. (II) is useful in forensic  
CC biology for detecting DNA sequences and as diagnostic probes for  
CC detecting the presence of specific mRNA in a particular cell type.  
XX  
SQ Sequence 1567 BP; 303 A; 510 C; 463 G; 291 T; 0 other;

Query Match 75.7%; Score 17.4; DB 21; Length 1567;

Best Local Similarity 94.7%; Pred. No. 71;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 atccatgttaagcagcagg 22  
||||||| |||||||  
Db 805 atccatgtgaagcagcagg 823

## RESULT 6

AAC39525

ID AAC39525 standard; DNA; 1153 BP.

XX AC AAC39525;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 24952.

XX KW Hybridisation assay; genetic mapping; gene expression control;

XX KW protein identification; signal transduction pathway;

XX KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 23-APR-1999; 99US-0130891.

XX PR 28-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 30-APR-1999; 99US-0132407.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 07-MAY-1999; 99US-0132487.

XX PR 11-MAY-1999; 99US-0132863.

XX PR 14-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.

```
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148884.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 18-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 74.8%; Score 17.2; DB 21; Length 1153;
Best Local Similarity 86.4%; Pred. No. 84;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaaacgacgagg 23
    |||||
Db 1088 aaatccatgtaaatcagtaagg 1109

RESULT 7
AAH18005
ID AAH18005 standard; cDNA; 2857 BP.
XX
AC AAH18005;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17809.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WPI; 2001-318749/34.
XX
Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
XX
Claim 8; SEQ ID 17809; 2537pp + CD ROM; English.
XX
The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
```

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2857 BP; 859 A; 564 C; 576 G; 858 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 2857;  
Best Local Similarity 86.4%; Pred. No. 96;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaagcagcagg 23  
|| ||| ||||| ||||| ||  
Db 1659 aactcctgtgtaagcagcag 1680

RESULT 8  
AAI58467  
ID AAI58467 standard; cDNA: 4495 BP.

XX AAI58467;

XX 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 670.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

OS WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM39511.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 670; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening, and  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

SQ Sequence 4495 BP; 1401 A; 945 C; 1040 G; 1109 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 4495;  
Best Local Similarity 86.4%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 aaatccatgtaagcagcagg 23  
|| ||| ||||| ||||| ||  
Db 3034 aactcctgtgtaagcagcag 3055

RESULT 9

AAI60253

ID AAI60253 standard; cDNA: 4775 BP.

XX AAI60253;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4242.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX Homo sapiens.

OS WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM41097.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4242; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC C.N.S disorders,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX SQ Sequence 4775 BP; 1494 A; 999 C; 1122 G; 1160 T; 0 other;

Query Match 74.8%; Score 17.2; DB 22; Length 4775;  
 Best Local Similarity 86.4%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgaggg 23  
 ||| ||||| ||||| |||||  
 Db 2995 aactcctgttaaacgacgagc 3016

RESULT 10  
 AAT74989/c  
 ID AAT74989 standard; DNA; 8826 BP.

XX AC AAT74989;

XX DT 26-FEB-1998 (first entry)

XX XX Saccharomyces cerevisiae YHK1243 encoding CSF-1 gene.

XX XX Saccharomyces cerevisiae; cold sensitivity of fermentability; CSF-1;  
 KW low temperature sensitivity; yeast; bread; ethanol; fermentation; ds.

XX OS Saccharomyces cerevisiae.

XX PN WO9724442-A1.

XX PD 10-JUL-1997.

XX PF 27-DEC-1996; 96WO-JP03862.

XX PR 28-DEC-1995; 95JP-0343700.

XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Kawasaki H, Kikuchi Y, Ouchi K, Tokai M;

XX WPI; 1997-363678/33.

DR P-PSDB; AAW22050.

XX PT Protein for complementing low temperature sensitivity in  
 PT fermentation - can be inactivated in yeast and used for producing  
 PT bread and ethanol

XX PS Claim 1; Pages 21-40; 48pp; Japanese.

XX CC This novel genomic DNA CSF-1 (cold sensitivity of fermentability) encodes  
 CC a protein capable of complementing a mutation showing low temperature  
 CC sensitivity in fermentation. The Saccharomyces cerevisiae YHK 1243 has an  
 CC inactivated gene encoding the new protein or has the novel DNA sequence  
 CC contained in its chromosome. The protein can be used in dough to produce  
 CC bread, and to produce ethanol.

XX SQ Sequence 8826 BP; 2903 A; 1551 C; 1760 G; 2612 T; 0 other;

Query Match 74.8%; Score 17.2; DB 18; Length 8826;

Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgaggg 23  
 ||| ||||| ||||| |||||  
 Db 7504 AAATTAATGTAAGCACCAGGG 7483

RESULT 11  
 AAT94548/c  
 ID AAT94548 standard; DNA; 9621 BP.

XX AC AAT94548;

XX DT 25-MAR-1998 (first entry)

XX XX Mutant YLR087c gene from cold sensitive yeast strain.

XX KW Mutant; cold sensitive; yeast; complementation; breadmaking; dough;  
 KW sugar; leavening agent; brewing; winemaking; truncation; ss.

XX OS Saccharomyces cerevisiae.

XX FH Key Location/Qualifiers

FT CDS 376..7866

FT /\*tag= a

FT /product= "mutant YLR087c protein"

XX PN WO9728693-A1.

XX PD 14-AUG-1997.

XX PF 07-FEB-1997; 97WO-FR00254.

XX PR 08-FEB-1996; 96FR-0001562.

XX PA (LESA ) LESAFFRE & CIE.

XX PI Colavizza D, Lolez A, Wadoux I;

XX DR WPI; 1997-414988/38.

DR P-PSDB; AAW36093.

XX PT New strains of bread-making yeast with low fermentative activity at  
 PT low temperature - allows production of doughs that can be stored  
 PT cold for many hours before final baking

XX PS Claim 10; Page 56-59; 76pp; French.

XX CC This is the nucleotide sequence of a mutant gene (designated YLR087c)  
 CC isolated from a cold sensitive strain of Saccharomyces cerevisiae. The  
 CC sequence was isolated from the clone YCP50-10.39 which was able to  
 CC complement the cold sensitive strain HLI3.2.30. This gene contains a  
 CC TCA to TGA mutation which encodes a truncated protein. The wild type  
 CC gene contains an open reading frame of 8874 bases encoding a protein of  
 CC 2958 amino acids, whereas this mutant gene contains an open reading frame  
 CC of 7488 bases encoding a protein of 2496 amino acids. Yeast strains  
 CC containing the YLR087c gene are used, fresh or dried, in modified  
 CC breadmaking processes, i.e. in processes where the delay between mixing  
 CC the dough and baking exceeds 6 hr. Particularly they are used to make  
 CC French-style bread (no added sugar) or breads with sugar content below  
 CC 5%. The strains can also be used to make mixed yeast/bacteria acidic  
 CC leavening agents. The strains can also be used in brewing and winemaking  
 CC as a cold-sensitive phenotype.

XX SQ Sequence 9621 BP; 3126 A; 1679 C; 1903 G; 2907 T; 6 other;

Query Match 74.8%; Score 17.2; DB 18; Length 9621;  
 Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgaggg 23

Db 7927 AATTTAAATGTAAGCACCAGG 7906  
||||| ||||||||| |||||

RESULT 12  
AAH29132/c

ID AAH29132 standard; DNA; 602 BP.

AC AAH29132;

DT 17-JUL-2001 (first entry)

XX Drosophila melanogaster essential gene fragment, SEQ ID NO: 321.

DE Drosophila melanogaster; fruit fly; essential gene; screening assay;

KW Drosophila melanogaster; fruit fly; essential gene; screening assay;

KW pesticide; crop protection; chromosome 3; ds.

XX Drosophila melanogaster.

OS WO200118547-A1.

PN 15-MAR-2001.

PD 06-SEP-2000; 2000WO-GB03444.

PF 07-SEP-1999; 99GB-0021009.

XX (UNIU ) UNIV GLASGOW.

PA Davies RW, Kaiser K, Yang MY;

PI WPI; 2001-281436/29.

DR Screening assays for used for identifying compounds having a

XX physiological effect on proteins identified as being essential -

PT Claim 1; Page 340-341; 695pp; English.

XX The present sequence is part of an essential gene from Drosophila

CC melanogaster. Lack of expression of the protein encoded by this

CC gene leads to a lethal or semi-lethal phenotype. The invention

CC relates to 902 nucleic acid sequences from genes encoding proteins

CC which are thought to be essential, and to a screening assay for

CC identifying compounds which have a physiological effect on these

CC proteins. Suitable compounds are useful as pesticides and may be used

CC in conjunction with other pesticides and herbicides for crop

CC protection. The gene corresponding to the present sequence is located

CC on chromosome 3.

XX Sequence 602 BP; 159 A; 131 C; 120 G; 192 T; 0 other;

Query Match 73.0%; Score 16.8; DB 22; Length 602;

Best Local Similarity 90.0%; Pred. No. 1.2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacag 21

Db 409 AGATCCATGTAATCAGCAG 390

RESULT 13

AAF57408/c

ID AAF57408 standard; DNA; 1752 BP.

XX AAF57408;

AC AAF57408;

XX 11-JUN-2001 (first entry)

DT Human p95 protein encoding DNA.

DE Rab; Rho; GTPase; pharmaceutical; cancer; anti-HIV; tuberculosstatic;

XX protozoacide; atidiabetic; nootropic; neuroprotective; dermatological;

KW

antipsoriatic; antiinflammatory; antiallergic; antipyretic; cytostatic;  
antibacterial; gynecological; antiatherosclerotic; gene therapy; human;  
p100; p95; p60; p45; p25; endocytic trafficking; GTPase effector; ds.

XX Homo sapiens.

OS Key Location/Qualifiers

XX CDS 1..1752

FT /\*tag= a

FT /product= "p95"

XX WO200120022-A1.

PN 22-MAR-2001.

PD 18-SEP-2000; 2000WO-EP09130.

PF 16-SEP-1999; 99EP-0118385.

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Nielsen E, Christophoridis S, Murphy C, Zerial M, De Renzis S;

PI WPI; 2001-257888/26.

XX P-PSDB; AAB62180.

DR Use of effectors of GTPase as target in a in vitro/vivo assay for

XX detecting substances for prophylaxis, treatment of cancer, cell

PT migration disorders, e.g. Alzheimer's, infectious diseases, diabetes,

PT atherosclerosis

XX Claim 10; Page 70-71; 76pp; English.

XX The invention relates to the use of effectors/regulators for Rab and Rho

CC GTPases in in vitro or in vivo assays. The method is useful for detecting

CC substances useful as pharmaceutical agents for the prophylaxis or

CC treatment of cancer and other proliferative, invasive or cell migration

CC disorders such as endometriosis, atherosclerosis, inflammatory and

CC allergic diseases, infectious diseases, diabetes, Alzheimer's disease and

CC skin repair diseases such as psoriasis. The infectious diseases include

CC AIDS, tuberculosis, pseudotuberculosis, cholera, gastroenteritis, enteric

CC fever, malaria, typhus, diseases caused by pathogens such as Listeria,

CC Mycobacterium, Staphylococcus, Toxoplasma, Trypanosoma, Salmonella,

CC Legionella, Leishmania, Coxiella, Shigella, Yersinia, Neisseria, Vibrio,

CC Bartonella. The cancer includes benign tumor, malignant lung carcinoma,

CC leukemia, glioma or a neuroblastoma, in particular lung carcinoma,

CC osteosarcoma, lymphoma, breast, bile, intestine, kidney, ovary, stomach,

CC brain, prostate, liver and every tumor that invades other tissues and

CC organs distinct from its site of origin. The assay is highly sensitive

CC and advantageous in the selectivity of the targets. The present sequence

CC represents the DNA encoding a human p95 protein. A multiprotein complex

CC including p100, p95, p60, p45, p25 (sequences AAF57404-408) acts as a

CC GTPase effector/regulator and has the function of endocytic trafficking.

XX Sequence 1752 BP; 496 A; 425 C; 444 G; 387 T; 0 other;

Query Match 73.0%; Score 16.8; DB 22; Length 1752;

Best Local Similarity 90.0%; Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gaatccatgtaaacgacga 20

Db 712 GAAATCCATGCAGACGACGA 693

RESULT 14

AAT60073

ID AAT60073 standard; DNA; 4388 BP.

XX AAT60073;

XX 13-MAY-1997 (first entry)

DT

XX Selective marker gene.  
DE  
XX  
XX Selective marker; enzyme: biosynthetic system; leucine synthesis;  
KW Rhizomucor pusillus; protein production; ss.  
XX  
XX Synthetic.  
OS  
XX JP09009971-A.  
PN  
XX 14-JAN-1997.  
PD  
XX  
XX 30-JUN-1995; 95JP-0166115.  
PF  
XX  
XX 30-JUN-1995; 95JP-0166115.  
PR  
XX  
XX (MEIT ) MEITO SANGYO KK.  
PA  
XX  
XX WPI; 1997-126428/12.  
DR  
XX  
XX Novel selective marker gene - used in a transformation system of  
PT Rhizomucor pusillus for protein production  
PT  
XX  
XX Claim 1; Page 9-10; 12pp; Japanese.  
PS  
XX  
XX This sequence represents the selective marker gene of the invention. This  
CC gene encodes an enzyme relating to the biosynthetic system of leucine  
CC synthesis. This sequence can be used in a recombinant vector for the  
CC transformation of Rhizomucor pusillus. The recombinant vector also  
CC includes another DNA sequence encoding a protein. The plasmid and the  
CC transformed Rhizomucor pusillus can then be used for the efficient  
CC production of useful proteins and peptides, which are encoded by the  
CC second DNA sequence in the vector.  
XX  
XX  
SQ Sequence 4388 BP; 1144 A; 1094 C; 1077 G; 1073 T; 0 other;

Query Match 73.0%; Score 16.8; DB 18; Length 4388;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagca 20  
||||| ||||| ||||| |||||  
Db 611 gaaatgcataaaagcagca 630

RESULT 15  
AAQ60834  
ID AAQ60834 standard; DNA; 281 BP.  
XX  
XX  
AC AAQ60834;  
XX  
DT 16-MAR-1994 (first entry)  
XX  
XX Human brain Expressed Sequence Tag EST00937.  
DE  
DE  
XX  
XX Gene transcription product; genetic markers; tagging; in vivo;  
KW transcription; mapping; locations; chromosomes; chromosomal; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO9316178-A.  
PN  
XX  
XX 19-AUG-1993.  
PD  
XX  
XX 12-FEB-1993; 93WO-US01294.  
PF  
XX  
XX 12-FEB-1992; 92US-0837195.  
PR  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PA  
XX  
XX Adams MD, Moreno RF, Venter CJ;  
PI  
XX

DR WPI; 1993-272882/34.  
XX  
XX Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes  
XX  
XX Example 4; Page 398; 500pp; English.  
PS  
XX  
XX The Expressed Sequence Tag was isolated from a human brain cDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes  
CC on chromosomes, for individual or forensic identification, for mapping  
CC locations of disease-associated genes, for identification of tissue  
CC type, and for prep. of antisense sequences, probes and constructs.  
CC EST00937 has a "poor" coding probability as evaluated using the  
XX coding-region prediction program CRM. See also AAQ59041-Q61440.  
XX  
SQ Sequence 281 BP; 45 A; 81 C; 91 G; 61 T; 3 other;

Query Match 72.2%; Score 16.6; DB 14; Length 281;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaagcagcagg 23  
||||| ||||| ||||| |||||  
Db 86 gaagtgaatgaaagcagcagg 108

Search completed: February 25, 2002, 18:17:40  
Job time: 16698 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:49 ; Search time 301.6 Seconds  
(without alignments)  
17.271 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23

Sequence: 1 gaatccatgtaaacgagcagg 23

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.2	74.8	8874	US-08-894-344C-1	Sequence 1, Appl
C 2	17.2	74.8	9621	US-09-125-028-1	Sequence 1, Appl
C 3	16.6	72.2	2460	US-08-964-127-1	Sequence 1, Appl
C 4	16.6	72.2	9751	US-09-238-303-7	Sequence 7, Appl
C 5	16.2	70.4	27	US-08-832-399-9	Sequence 9, Appl
C 6	16.2	70.4	27	US-09-372-498-9	Sequence 9, Appl
C 7	16.2	70.4	717	US-08-706-702-25	Sequence 25, Appl
C 8	16.2	70.4	717	US-08-706-706-25	Sequence 25, Appl
C 9	16.2	70.4	1351	US-09-002-114-2	Sequence 2, Appl
C 10	16.2	70.4	1608	US-08-899-371-2	Sequence 2, Appl
C 11	16.2	70.4	1613	US-08-899-371-1	Sequence 1, Appl
C 12	16.2	70.4	5183	US-09-039-555B-18	Sequence 18, Appl
C 13	16.2	70.4	5243	US-08-414-335-2	Sequence 2, Appl
C 14	16.2	70.4	16442	US-08-781-891-208	Sequence 208, Appl
C 15	15.8	68.7	199	US-08-250-346-21	Sequence 21, Appl
C 16	15.8	68.7	394	US-07-885-089B-15	Sequence 15, Appl
C 17	15.8	68.7	394	US-07-885-089B-17	Sequence 17, Appl
C 18	15.8	68.7	3410	US-09-020-956-110	Sequence 110, Appl
C 19	15.8	68.7	3410	US-09-030-607-110	Sequence 110, Appl
C 20	15.8	68.7	3466	US-08-468-036-38	Sequence 38, Appl
C 21	15.8	68.7	3466	US-08-376-843-38	Sequence 38, Appl
C 22	15.6	67.8	1098	US-08-948-616-6	Sequence 6, Appl
C 23	15.6	67.8	1098	US-09-193-510-6	Sequence 6, Appl
C 24	15.6	67.8	1098	US-09-368-402-6	Sequence 6, Appl
C 25	15.6	67.8	1488	US-09-171-969-1	Sequence 1, Appl
C 26	15.4	67.0	2266	US-09-213-767-1	Sequence 1, Appl
C 27	15.4	67.0	152331	US-09-128-155-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1  
US-08-894-344C-1/c  
; Sequence 1, Application US/08894344C  
; Patent No. 6172196  
; GENERAL INFORMATION:  
; APPLICANT: KAWASAKI, Hideki  
; APPLICANT: TOKAI, Masaya  
; APPLICANT: KIKUCHI, Yasuhiro  
; APPLICANT: OUCHI, Koza  
; TITLE OF INVENTION: DNA ENCODING PROTEIN COMPLEMENTING  
; TITLE OF INVENTION: YEAST  
; TITLE OF INVENTION: LOW TEMPERATURE-SENSITIVE FERMENTABILITY  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FITZPATRICK, CELLA, HARPER & SCINTO  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10112-3801  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette - 3.50 inch, 1440 Kb storage.  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS Ver3.30  
; SOFTWARE: PATENT AID Ver1.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/894,344C  
; FILING DATE: 15-AUGUST-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP343700/95  
; FILING DATE: 28-DECEMBER-1995  
; APPLICATION NUMBER: PCT/JP96/03862  
; FILING DATE: 27-DECEMBER-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perry, Lawrence S.  
; REGISTRATION NUMBER: 31865  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-218-2100  
; TELEFAX: 212-218-2200  
; INFORMATION FOR SEQ ID NO: 1 :  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8874 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE: *Saccharomyces cerevisiae*  
; ORGANISM: *Saccharomyces cerevisiae*  
; STRAIN: X2180-1B  
; FEATURE:

C 28	15.4	67.0	176373	3	US-09-128-155-17	Sequence 17, Appl
C 29	15.2	66.1	681	5	PCT-US94-04361-22	Sequence 22, Appl
C 30	15.2	66.1	1338	2	US-08-484-993B-7	Sequence 7, Appl
C 31	15.2	66.1	1338	2	US-08-484-158B-7	Sequence 7, Appl
C 32	15.2	66.1	1338	2	US-08-484-596A-7	Sequence 7, Appl
C 33	15.2	66.1	1338	2	US-08-480-150A-7	Sequence 7, Appl
C 34	15.2	66.1	1338	3	US-08-458-731-7	Sequence 7, Appl
C 35	15.2	66.1	1338	3	US-08-149-223A-7	Sequence 7, Appl
C 36	15.2	66.1	2975	4	US-08-368-281-1	Sequence 3, Appl
C 37	15.2	66.1	5107	4	US-08-910-647-3	Sequence 3, Appl
C 38	15.2	66.1	18994	1	US-08-459-586-4	Sequence 4, Appl
C 39	15.2	66.1	18994	2	US-08-282-696-4	Sequence 4, Appl
C 40	15	65.2	35	1	US-08-094-534-18	Sequence 18, Appl
C 41	15	65.2	35	2	US-08-581-543-18	Sequence 18, Appl
C 42	15	65.2	35	5	PCT-US94-08000-18	Sequence 18, Appl
C 43	15	65.2	39	1	US-08-094-534-17	Sequence 17, Appl
C 44	15	65.2	39	2	US-08-581-543-17	Sequence 17, Appl
C 45	15	65.2	39	5	PCT-US94-08000-17	Sequence 17, Appl

```
; NAME/KEY: CDS
; LOCATION: 1 to 8874
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 1291 to 1296
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 4388 to 4393
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 5927 to 5032
; IDENTIFICATION METHOD: S
; FEATURE:
; NAME/KEY: cleavage-site
; LOCATION: 7675 to 7680
; IDENTIFICATION METHOD: S
; US-08-894-344C-1
```

```
Query Match 74.8%; Score 17.2; DB 4; Length 8874;
Best Local Similarity 86.4%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 aaatccatgtaaacgacgagg 23
||||| ||||||| |||||
Db 7552 AAATTAATGTAAGCACCAGG 7531
```

## RESULT 2

```
US-09-125-028-1/c
; Sequence 1, Application US/09125028A
; Patent No. 6190707
; GENERAL INFORMATION:
; APPLICANT: WADOUX, Isabelle
; APPLICANT: COLAVIZZA, Didier
; TITLE OF INVENTION: NOVEL COLD-SENSITIVE BREAD-MAKING YEASTS
; FILE REFERENCE: levure sensible froid
; CURRENT APPLICATION NUMBER: US/09/125,028A
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: PCT/FR97/00254
; EARLIER FILING DATE: 1997-02-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 9621
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9318)
; OTHER INFORMATION: Use of n signifies any of g, a, c or t
; US-09-125-028-1
```

```
Query Match 74.8%; Score 17.2; DB 4; Length 9621;
Best Local Similarity 86.4%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 aaatccatgtaaacgacgagg 23
||||| ||||||| |||||
Db 7927 AAATTAATGTAAGCACCAGG 7906
```

## RESULT 3

```
US-08-964-127-1/c
; Sequence 1, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
```

```
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
; REFERENCE/DOCKET NUMBER: 07334/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 498....2057
; US-08-964-127-1
```

```
Query Match 72.2%; Score 16.6; DB 4; Length 2460;
Best Local Similarity 82.6%; Pred. No. 41;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 gaaatccatgtaaacgacgagg 23
||||| ||||||| |||||
Db 2242 GAAGTGAATGCAAGCAGCAGG 2220
```

## RESULT 4

```
US-09-238-303-7
; Sequence 7, Application US/09238303B
; Patent No. 6284253
; GENERAL INFORMATION:
; APPLICANT: Barr, Margaret C.
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequence
; FILE REFERENCE: 18617,0059
; CURRENT APPLICATION NUMBER: US/09/238,303B
; CURRENT FILING DATE: 1999-01-28
; EARLIER APPLICATION NUMBER: US 60/072,927
; EARLIER FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 7
; LENGTH: 9751
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant viral clone constructed from the genomic DNA of
; OTHER INFORMATION: a Pallas's cat feline immunodeficiency virus
; US-09-238-303-7
```

```

; FILE OF INVENTION: GH-50020-1
; CURRENT APPLICATION NUMBER: US/09/372,498
; CURRENT FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: 08/832,399
; PRIOR FILING DATE: 1997-04-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 27
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
; US-09-372-498-9

Query Match 70.4%; Score 16.2; DB 4; Length 27;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 3; Indels

QY 1 gaatccatgtaagcagcag 21
   ||| ||||| ||| |||||
Db 6 gaattccatggaaccagcag 26

RESULT 7
US-08-706-702-25/c
; Sequence 25, Application US/08706702
; Patent No. 5948614
; GENERAL INFORMATION:
; APPLICANT: Chatterjee, Deb K.
; TITLE OF INVENTION: Cloned DNA Polymerases from Thermotoga
; TITLE OF INVENTION: maritima and Mutants Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,702
; FILING DATE: 06-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/689,807
; FILING DATE: 14-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,400
; FILING DATE: 02-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,759
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/537,397
; FILING DATE: 02-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/525,057
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REFERENCE/DOCKET NUMBER: 32,893
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:

```



```
; Sequence 2, Application US/08899371
; Patent No. 5962227
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Ronald P.
; APPLICANT: Andree, Karl B.
; APPLICANT: Antonio, Dolores B.
; TITLE OF INVENTION: A DNA-Based Diagnostic Test for
; TITLE OF INVENTION: Detecting Myxobolus, the Cause of Salmonid Whirling
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsens and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,371
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,734
; FILING DATE: 26-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-081310US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1608 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1608
; OTHER INFORMATION: /note= "18S rRNA gene of Myxobolus
; insidiosus"
;
US-08-899-371-2

Query Match 70.4%; Score 16.2; DB 2; Length 1608;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaagcagcagg 22
| | | | | | | | | | | | | | | |
Db 312 ACATCCATGGAAGCAGCAGG 332

RESULT 11
US-08-899-371-1
; Sequence 1, Application US/08899371
; Patent No. 5962227
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Ronald P.
; APPLICANT: Andree, Karl B.
; APPLICANT: Antonio, Dolores B.
; TITLE OF INVENTION: A DNA-Based Diagnostic Test for
; TITLE OF INVENTION: Detecting Myxobolus, the Cause of Salmonid Whirling
; TITLE OF INVENTION: Disease
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Townsens and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,371
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/022,734
; FILING DATE: 26-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-081310US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..1613
; OTHER INFORMATION: /note= "18S rRNA gene of Myxobolus
; cerebrealis"
;
US-08-899-371-1

Query Match 70.4%; Score 16.2; DB 2; Length 1613;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaagcagcagg 22
| | | | | | | | | | | | | | | |
Db 315 ACATCCATGGAAGCAGCAGG 335

RESULT 12
US-09-039-555B-18/C
; Sequence 18, Application US/09039555B
; Patent No. 6033856
; GENERAL INFORMATION:
; APPLICANT: Koerner, Kathrin
; APPLICANT: Mueller, Rolf
; APPLICANT: Sadlacek, Hans-Harald
; TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/09/039,555B  
FILING DATE: 16-MAR-1998  
CLASSIFICATION: 514  
PRIOR APPLICATION NUMBER: DE 19710643.9  
FILING DATE: 14-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 016779/0131  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5183 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-039-555B-18

Query Match 70.4%; Score 16.2; DB 3; Length 5183;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgag 22  
||||||| |||||||||  
Db 741 AAATCCAGCTATAGCAGCAGG 721

RESULT 13  
US-08-414-335-2  
Sequence 2, Application US/08414335  
Patent No. 5907078  
GENERAL INFORMATION:  
APPLICANT: GREENBERT, No. 5907078man M  
APPLICANT: MATUSIK, Robert J  
APPLICANT: ROSEN, Jeffrey M  
TITLE OF INVENTION: TRANSGENIC MOUSE MODEL FOR PROSTATE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare Ltd.  
STREET: 1203 Crystal Plaza Bldg. I, 2001 Jefferson  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/414,335  
FILING DATE: 31-MAR-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: FALLOW, Charles W  
REGISTRATION NUMBER: 28,946  
REFERENCE/DOCKET NUMBER: 1027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
US-08-414-335-2

Query Match 70.4%; Score 16.2; DB 2; Length 5243;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaacgacgag 22  
||||||| |||||||||  
Db 4443 AAATCCAGCTATAGCAGCAGG 4463

RESULT 14  
US-08-781-891-208/c  
Sequence 208, Application US/08781891  
Patent No. 6090620  
GENERAL INFORMATION:  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Yu, Chang-En  
APPLICANT: Oshima, Junko  
APPLICANT: Mulligan, John T.  
APPLICANT: Schellenberg, Gerald D.  
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,891  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6090620tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 240052.419  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16442 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-781-891-208

Query Match 70.4%; Score 16.2; DB 3; Length 16442;  
Best Local Similarity 85.7%; Pred. No. 94;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaatccatgtaaacgacgag 21  
||||||| |||||||||  
Db 12904 GAAATCCATGTAAGTCACAG 12884

RESULT 15  
US-08-250-346-21/c  
Sequence 21, Application US/08250346  
Patent No. 5939255  
GENERAL INFORMATION:  
APPLICANT: ANAND, Rakesh

APPLICANT: MARKHAM, Alexander F  
APPLICANT: SMITH, John C  
APPLICANT: ANWAR, Rashida  
APPLICANT: RILEY, John H  
APPLICANT: OGILVIE, Donald J  
APPLICANT: ELVIN, Paul  
TITLE OF INVENTION: DIAGNOSTIC METHOD  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/250,346  
FILING DATE: 27-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/115,675  
FILING DATE: 02-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/879,117  
FILING DATE: 04-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/578,616  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8920211.3  
FILING DATE: 07-SEP-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: BIRD, Donald J.  
REGISTRATION NUMBER: 25,323  
REFERENCE/DOCKET NUMBER: 205356/PHM35417/USC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 199 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-250-346-21

Query Match 68.7%; Score 15.8; DB 2; Length 199;  
Best Local Similarity 89.5%; Pred. No. 60;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 tccatgtaaagcagcagg 23  
|||||  
Db 178 TCCAGGAAAGCAGCAGG 160

Search completed: February 25, 2002, 18:05:51  
Job time: 18599 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:13 ; Search time 8261.74 seconds  
(without alignments)  
29.915 Million cell updates/sec

Title: US-09-698-903B-11

Perfect score: 23  
Sequence: 1 gaatccatgttaagcagcagg 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estfun:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estom:\*
- 5: em\_estpl:\*
- 6: em\_estba:\*
- 7: em\_estro:\*
- 8: em\_estov:\*
- 9: em\_hic:\*
- 10: gb\_estl:\*
- 11: gb\_est2:\*
- 12: gb\_hic:\*
- 13: gb\_gss:\*
- 14: em\_gss\_fun:\*
- 15: em\_gss\_hum:\*
- 16: em\_gss\_inv:\*
- 17: em\_gss\_pln:\*
- 18: em\_gss\_pro:\*
- 19: em\_gss\_rod:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.8	81.7	619	13	FR0032239
2	18.8	81.7	916	13	AL028607 Fugu rubr
3	18.8	81.7	1007	13	AL315856 Tetraodon
C 4	18.4	80.0	342	13	AL345585 Tetraodon
C 5	18.4	80.0	522	13	AZ842531
6	18.4	80.0	618	13	AQ972989 Fugu rubr
C 7	18.2	79.1	205	10	FR0034498
C 8	18.2	79.1	287	10	AV119745
C 9	18.2	79.1	444	10	AV119723
10	18.2	79.1	455	13	AI243356
C 11	18.2	79.1	481	10	AZ628185
C 12	18.2	79.1	533	13	BE019857
					AZ484979

C 13	18.2	79.1	750	13	AZ750976
C 14	18.2	79.1	809	11	BF027105
C 15	18.2	79.1	875	11	BF027105
C 16	18.2	79.1	951	13	BG118586
C 17	17.8	77.4	193	13	CNS0528P
C 18	17.8	77.4	193	13	CNS0528P
C 19	17.8	77.4	287	10	AZ658541
C 20	17.8	77.4	525	10	BE579037
C 21	17.8	77.4	563	10	AV395870
C 22	17.8	77.4	564	10	AV395870
C 23	17.8	77.4	582	13	AV387535
C 24	17.8	77.4	595	10	AV387535
C 25	17.8	77.4	618	10	AZ107460
C 26	17.8	77.4	626	10	RPCI-23-3
C 27	17.8	77.4	630	10	AA394343
C 28	17.8	77.4	657	10	AW661616
C 29	17.8	77.4	700	11	AW661616
C 30	17.8	77.4	723	11	AW661616
C 31	17.8	77.4	723	11	AW661616
C 32	17.4	75.7	221	10	AZ558735
C 33	17.4	75.7	364	10	BE337492
C 34	17.4	75.7	604	13	BE238191
C 35	17.4	75.7	857	13	BE238191
C 36	17.4	75.7	975	11	AW054024
C 37	17.4	75.7	1011	13	AW054024
C 38	17.2	74.8	158	11	EG846208
C 39	17.2	74.8	226	10	EG846208
C 40	17.2	74.8	271	10	AA298276
C 41	17.2	74.8	312	10	AA298276
C 42	17.2	74.8	312	11	AW357388
C 43	17.2	74.8	332	10	AI905219
C 44	17.2	74.8	340	10	H20500
C 45	17.2	74.8	357	11	BB251925
					AA012198
					W63225

## ALIGNMENTS

RESULT 1	FR0032239	619 bp	DNA	GSS	27-JUN-1998
LOCUS	Fugu rubripes	GSS sequence, clone 137L16aC10,	genomic survey		
DEFINITION	Sequence.				
ACCESSION	AL028607				
VERSION	AL028607.1	GI:3270721			
KEYWORDS	GSS: genome survey sequence.				
SOURCE	Takifugu rubripes.				
ORGANISM	Takifugu rubripes				
REFERENCE	1 (bases 1 to 619)				
AUTHORS	Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrana,Y., Williams,G. and Brenner,S.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK. Email: bhoelp@hmp.mrc.ac.uk				
COMMENT	Vector: pBluescript II KS				
	V.type: phagemid				
	PRIMER: KS				
DESCR:	One pass dye-terminator sequencing of cosmid cloned genomic sequence.				
FEATURES	Location/Qualifiers				
source	1..619				
	/organism="Takifugu rubripes"				
	/db_xref="taxon:31033"				
	/clone_lib="cosmid_137L16"				
	/clone="137L16aC10"				
BASE COUNT	131 a 178 c 145 g 116 t 49 others				
ORIGIN					



```

JOURNAL COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: gdunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0141 row: J column: 03
Seq primer: CGTGTAAAGCAGCGCCAGT
Class: plasmid ends
High quality sequence stop: 342.
Location/Qualifiers
1. 342
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M014J03"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 50 a 84 c 82 g 126 t
ORIGIN

Query Match 80.0%; Score 18.4; DB 13; Length 342;
Best Local Similarity 95.0%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 atccatgtaaagcagcagg 23
| ||||| ||||| |||||
Db 174 ACCCATGTAAAGCAGCAGG 155

RESULT 5
LOCUS AO972989/c
DEFINITION AO972989.1 GI:6803442
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 522)
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akınret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P., and Fraser, C.M.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)

Other_GSSs: RPCI-23-317G14.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 317 row: G column: 14
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 522
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-317G14"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1; EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 90 a 159 c 106 g 164 t 3 others
ORIGIN

Query Match 80.0%; Score 18.4; DB 13; Length 522;
Best Local Similarity 95.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 atccatgtaaagcagcagg 23
| ||||| ||||| |||||
Db 394 ACCCATGTAAAGCAGCAGG 375

RESULT 6
LOCUS FR0034498
DEFINITION Fugu rubripes GSS sequence, clone 199F09af9, genomic survey sequence.
ACCESSION AL030865
VERSION AL030865.1 GI:3272979
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 618)
AUTHORS Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrantia, Y., Williams, G. and Brenner, S.
Direct Submission
TITLE Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK. Email: biohelp@hmp.mrc.ac.uk
JOURNAL Vector: pBluescript II KS
COMMENT V_type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.

```

```

FEATURES
  source
    Location/Qualifiers
      1..618
        /organism="Takifugu rubripes"
        /db_xref="taxon:31033"
        /clone_lib="cosmid 199F09"
        /clone="199F09aF9"
BASE COUNT      109 a 183 c 154 g 141 t 31 others
ORIGIN

Query Match      80.0%; Score 18.4; DB 13; Length 618;
Best Local Similarity 86.4%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 aaatccatgtaaagcagcagg 23
Db 362 ACATCCNNGTAAGCAGCAGG 383

RESULT 7
AV119745/c
LOCUS      AV119745 Mus musculus C57BL/6J 10-day embryo Mus musculus CDNA
DEFINITION
ACCESSION  AV119745
VERSION     AV119745
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE    1 (bases 1 to 205)
AUTHORS      Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
              Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
              A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
              Kikuchi,N., Kojima,Y., Matsuyama,T., Naitsuma,H., Oda,H., Owa,C.,
              Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
              Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomimaga,N.,
              Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
              Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
              RIKEN Mouse ESTs
              Unpublished (1999)
              Contact: Chie Owa
              Genome Science Laboratory
              RIKEN
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: 81-298-36-9145
              Fax: 81-298-36-9098
              Email: genome-res@rtc.riken.go.jp
              Thermotabilization and thermoactivation of thermostable enzymes by
              trehalose and its application for the synthesis of full length cDNA
              (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
              Transcriptional sequencing: A method for DNA sequencing using RNA
              polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
              Please visit our web site (http://genome.rtc.riken.go.jp) for
              further details.
              Location/Qualifiers
                1..205
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="2610305L08"
                  /clone_lib="Mus musculus C57BL/6J 10-day embryo"
                  /sex="mixed"
                  /dev_stage="10-day embryo"
BASE COUNT      42 a 36 c 37 g 90 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 10; Length 205;
Best Local Similarity 87.0%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

FEATURES
  source
    Location/Qualifiers
      1..205
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="2610305L08"
        /clone_lib="Mus musculus C57BL/6J 10-day embryo"
        /sex="mixed"
        /dev_stage="10-day embryo"
BASE COUNT      42 a 36 c 37 g 90 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 10; Length 205;
Best Local Similarity 87.0%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

FEATURES
  source
    Location/Qualifiers
      1..287
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="2610305J08"
        /clone_lib="Mus musculus C57BL/6J 10-day embryo"
        /sex="mixed"
        /dev_stage="10-day embryo"
BASE COUNT      66 a 57 c 52 g 112 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 10; Length 287;
Best Local Similarity 87.0%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaaatccatgtaaagcagcagg 23
Db 192 GAAATCCAAGTAAACAGAGGG 170

RESULT 8
AV119723/c
LOCUS      AV119723 Mus musculus C57BL/6J 10-day embryo Mus musculus CDNA
DEFINITION
ACCESSION  AV119723
VERSION     AV119723
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE    1 (bases 1 to 287)
AUTHORS      Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
              Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
              A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
              Kikuchi,N., Kojima,Y., Matsuyama,T., Naitsuma,H., Oda,H., Owa,C.,
              Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara
              Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomimaga,N.,
              Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
              Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
              RIKEN Mouse ESTs
              Unpublished (1999)
              Contact: Chie Owa
              Genome Science Laboratory
              RIKEN
              3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
              Tel: 81-298-36-9145
              Fax: 81-298-36-9098
              Email: genome-res@rtc.riken.go.jp
              Thermotabilization and thermoactivation of thermostable enzymes by
              trehalose and its application for the synthesis of full length cDNA
              (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
              Transcriptional sequencing: A method for DNA sequencing using RNA
              polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
              Please visit our web site (http://genome.rtc.riken.go.jp) for
              further details.
              Location/Qualifiers
                1..287
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="2610305J08"
                  /clone_lib="Mus musculus C57BL/6J 10-day embryo"
                  /sex="mixed"
                  /dev_stage="10-day embryo"
BASE COUNT      66 a 57 c 52 g 112 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 10; Length 287;
Best Local Similarity 87.0%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gaaatccatgtaaagcagcagg 23
Db 274 GAAATCCAAGTAAACAGAGGG 252

RESULT 9
AI243356/c
LOCUS      AI243356 Homo sapiens CDNA clone
DEFINITION  q130902.x1 Soares_NFL_T_GBC_S1 Homo sapiens CDNA clone
ACCESSION  AI243356
VERSION     AI243356
KEYWORDS    EST.
SOURCE      human.

```



image.llnl.gov/image/html/iresources.shtml

Trace considered overall poor quality

Seq primer: -40RP from Gibco

High quality sequence stop: 1.

Location/Qualifiers

#### FEATURES

source

```
1. .481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3029912"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT

ORIGIN

117 a 127 c 138 g 99 t

Query Match 79.1%; Score 18.2; DB 10; Length 481;

Best Local Similarity 87.0%; Pred. No. 4.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgaggg 23

||||| ||||| ||||| ||||| |||||

Db 259 GAAGTCCATGTAGAGCAACAGGG 237

RESULT 12

AZ484979/c

LOCUS

DEFINITION 1M0311P16R mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0311P16 R, DNA sequence.

ACCESSTON AZ484979

VERSION 1 GI:10650348

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 533)

REFERENCE Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D.,Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 306, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Place: 0311 row: P column: 16

Seq primer: CACACGAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 533.

Location/Qualifiers

1. 533

FEATURES

source

```
/organism="Mus musculus"
/db_xref="taxon:9606"
/clone="UUGC1M0311P16"
```

```
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

```

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adaptor mouse DNA was annealed to

adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 139 a 119 c 103 g 172 t

ORIGIN

Query Match 79.1%; Score 18.2; DB 13; Length 533;

Best Local Similarity 87.0%; Pred. No. 4.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gaaatccatgtaaacgacgaggg 23

||||| ||||| ||||| ||||| |||||

Db 146 GAAGTCTATGTAAAGCAGCAGAG 124

RESULT 13

AZ750976/c

LOCUS

DEFINITION

RPci-24-128J23-TV RPci-24

Mus musculus genomic clone

RPci-24-128J23

, DNA sequence.

ACCESSTON AZ750976

VERSION AZ750976.1

KEYWORDS GI:12536135

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 750)

REFERENCE Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPci-24

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPci-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 128 row: J column: 23

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. 750

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

[illegible]

Qy	1	gaaatccaatgaagcagcgagg	23
Db	586	GAAACCCCTGCAAAAGCAGCGGG	608

  

RESULT	15
BGI18586/c	
LOCUS	BGI18586.1 875 bp mRNA EST 30-JAN-2001
DEFINITION	602348330F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4443391 5';
ACCSSION	mRNA sequence.
VERSION	BGI18586
KEYWORDS	BGI18586.1 GI:12612092
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (bases 1 to 875)
COMMENT	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM10217 row: k column: 08 High quality sequence stop: 581. Location/Qualifiers 1..875 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4443391" /clone_lib="NIH_MGC_90" /tissue_type="adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: liver; Vector: pCMV-Sport6; Site:1; Noti: Average_2; Sali; Cloned unidirectionally; oligo-dT primed. Size insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC Library." Base COUNT 207 a 256 c 254 g 158 t ORIGIN

  

Query Match	79.1%	Score 18.2:	DB 11;	Length 875;
Best Local Similarity	87.0%	Pred. No. 5e+02;		
Matches 20:	Conservative	0:	Mismatches	3:
			Indels	0:
			Gaps	

Search completed: February 25, 2002, 17:21:17  
Job time: 16170 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:02:56 ; Search time 2331.3 Seconds  
(without alignments)  
148.604 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataacttgac 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Match	Length	DB	ID	Description
1	21	100.0	21	6	AX127759	AX127759 Sequence
2	19.4	92.4	21	6	AX172451	AX172451 Sequence
3	19.4	92.4	249	12	ARGMTUB	X05579 Soybean bet
c 4	19.4	92.4	415	6	AX127755	AX127755 Sequence
5	19.4	92.4	416	6	AX127757	AX127757 Sequence
c 6	19.4	92.4	694	6	AX172478	AX172478 Sequence
7	19.4	92.4	831	1	ATTRN7	V00090 Agrobacteri
8	19.4	92.4	878	1	ATTDNA	X00431 Agrobacteri
c 9	19.4	92.4	909	6	AX172475	AX172475 Sequence
10	19.4	92.4	1037	6	A10942	A10942 Nucleotide
c 11	19.4	92.4	1077	6	AX172463	AX172463 Sequence
12	19.4	92.4	1085	6	A10939	A10939 Nucleotide
13	19.4	92.4	1160	6	A10943	A10943 Nucleotide
14	19.4	92.4	1186	6	A18051	A18051 DNA used as
15	19.4	92.4	1186	6	AR095107	AR095107 Sequence
16	19.4	92.4	1186	6	AR098313	AR098313 Sequence
17	19.4	92.4	1186	6	AX012338	AX012338 Sequence
18	19.4	92.4	1186	6	I49886	I49886 Sequence 2
19	19.4	92.4	1186	6	I82374	I82374 Sequence 2
c 20	19.4	92.4	2476	12	TBI251013	AJ251013 Transform
21	19.4	92.4	3200	6	I44104	I44104 Sequence 23
22	19.4	92.4	3201	6	I44103	I44103 Sequence 22
c 23	19.4	92.4	3236	12	TBI251014	AJ251014 Transform
c 24	19.4	92.4	4832	6	AX172441	AX172441 Sequence
c 25	19.4	92.4	4946	6	A60108	A60108 Sequence 1
c 26	19.4	92.4	4946	6	A76915	A76915 Sequence 1
c 27	19.4	92.4	4946	6	AR098307	AR098307 Sequence
c 28	19.4	92.4	4946	6	AX172440	AX172440 Sequence 7
c 29	19.4	92.4	5349	6	A71437	A71437 Sequence 7
c 30	19.4	92.4	5560	6	A60112	A60112 Sequence 5
c 31	19.4	92.4	5560	6	AR098311	AR098311 Sequence
c 32	19.4	92.4	5865	6	AX127748	AX127748 Sequence
c 33	19.4	92.4	5865	6	AX127748	AX127748 Sequence
34	19.4	92.4	6539	6	E31991	E31991 Mutated bar
35	19.4	92.4	6548	6	A60109	A60109 Sequence 2
36	19.4	92.4	6548	6	A76916	A76916 Sequence 2
37	19.4	92.4	6548	6	AR098308	AR098308 Sequence
38	19.4	92.4	6548	6	E31990	E31990 Mutated bar
39	19.4	92.4	7566	6	A24783	A24783 plasmid pPS
40	19.4	92.4	7566	6	AR074388	AR074388 Sequence
41	19.4	92.4	7599	6	AX063413	AX063413 Sequence
42	19.4	92.4	7639	6	A24782	A24782 plasmid pJD
43	19.4	92.4	7639	6	AR074387	AR074387 Sequence
c 44	19.4	92.4	7811	6	AR078675	AR078675 Sequence
45	19.4	92.4	12095	12	BINHYGDA	Z37515 Binary vect

ALIGNMENTS

RESULT 1	AX127759	AX127759	21 bp	DNA	PAT	15-MAY-2001
LOCUS	AX127759	Sequence 12 from Patent WO0131042.				
DEFINITION	AX127759					
ACCESSION	AX127759.1	GI:14134406				
VERSION	AX127759.1					
KEYWORDS		synthetic construct.				
SOURCE		synthetic construct				
ORGANISM		artificial sequence.				
REFERENCE		1 (bases 1 to 21)				
AUTHORS		Weston,B. and de Beuckeleer,M.				
TITLE		Male-sterile brassica plants and methods for producing same				
JOURNAL		Patent: WO 0131042-A 12 03-MAY-2001;				
FEATURES		Avantis CropScience N.V. (BE)				
source		Location/Qualifiers				
		1..21				
		/organism="synthetic construct"				
		/db_xref="taxon:32630"				
		/note="primer M8201"				



```
ORGANISM    synthetic construct
            artificial sequence.
REFERENCE   1 (bases 1 to 416)
AUTHORS    Weston,B. and de Beuckeleer,M.
TITLE      Male-sterile brassica plants and methods for producing same
JOURNAL    Patent: WO 0131042-A 10 03-MAY-2001;
            Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
              1..416
              /organism="synthetic construct"
              /db_xref="taxon:32630"
              /note="3' border flanking region of elite event MS-B2"
            misc_feature
              1..193
              /note="T-DNA"
            misc_feature
              194..416
              /note="plant DNA"
BASE COUNT 137 a 72 c 54 g 152 t 1 others
ORIGIN
            1 gcttgactataacttgac 21
            |||||||||||||||
Db          81 GCTTGGACTATAATACCTGAC 101

Query Match      92.4%; Score 19.4; DB 6; Length 416;
Best Local Similarity 95.2%; Pred. NO. 12;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21
    |||||||||||||||
Db 81 GCTTGGACTATAATACCTGAC 101

RESULT 6
LOCUS      AX172478 694 bp DNA
DEFINITION Sequence 39 from Patent WO0141558.
ACCESSION  AX172478
VERSION    AX172478.1 GI:14597590
KEYWORDS   synthetic construct.
SOURCE     synthetic construct.
ORGANISM   synthetic construct.
            artificial sequence.
REFERENCE   1 (bases 1 to 694)
AUTHORS    de Both,G. and de Beuckeleer,M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL    Patent: WO 0141558-A 39 14-JUN-2001;
            Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
              1..694
              /organism="synthetic construct"
              /db_xref="taxon:32630"
              /note="sequence comprising the 5' flanking region of
              RF-BN1 in WOSR"
BASE COUNT 298 a 81 c 73 g 242 t

Query Match      92.4%; Score 19.4; DB 6; Length 694;
Best Local Similarity 95.2%; Pred. NO. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21
    |||||||||||||||
Db 694 GCTTGGACTATAATACCTGAC 674

RESULT 7
LOCUS      ATFRN7 831 bp DNA
DEFINITION Agrobacterium tumefaciens gene encoding 'transcript 7' mRNA for a
            protein with unknown function.
ACCESSION  V00090
VERSION    V00090.1 GI:39180
KEYWORDS   unidentified reading frame.
SOURCE     Agrobacterium tumefaciens.
ORGANISM   Agrobacterium tumefaciens.

Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group:
Rhizobiaceae; Rhizobium.
1 (bases 1 to 831)
Dhaese,P., De Greve,H., Gielen,J., Seurinck,J., Van Montagu,M.M.
and Schell,J.
Identification of sequences involved in the polyadenylation of
higher plant nuclear transcripts using Agrobacterium T-DNA genes as
models
EMBO J. 2, 419-426 (1983)
2 (bases 76 to 100)
Dhaese,P.
Direct Submission
Submitted (27-MAY-1983) to the EMBL/GenBank/DBJ databases
Data kindly reviewed (27-MAY-1983) by Dhaese P.
Location/Qualifiers
source
  1..831
  /organism="Agrobacterium tumefaciens"
  /strain="(octopine TL-DNA)"
  /db_xref="taxon:358"
  75..81
  /note="CG [1] revised CCAGAGG [2]"
  /citation=[1]
  /citation=[2]
  99..101
  /note="CCA [1] revised CTA [2]"
  /citation=[1]
  /citation=[2]
  132..673
  /note="transcript 7 (alternate)"
  132..672
  /note="transcript 7"
  148..528
  /note="unknown gene (148 is 1st base in codon) (525 is 3rd
  base in codon)"
  /codon_start=1
  /transl_table=11
  /protein_id="CAA23429.1"
  /db_xref="GI:39181"
  /db_xref="SWISS-PROT:P03867"
  /translation="MNFADTPLASLDLDNACEEFIKTYGASPOLETGCVIOTNNGLLY
  LYKGSLSORIHDTHLAFKEKEELSFTTIKPAEMKAQSDTTYVAIQSNFLCVSN
  PEKGLRCHNRPFILYPIVAHGSM"
BASE COUNT 262 a 176 c 138 g 255 t
ORIGIN
            1 gcttgactataacttgac 21
            |||||||||||||||
            664 GCTTGGACTATAATACCTGAC 684

Query Match      92.4%; Score 19.4; DB 1; Length 831;
Best Local Similarity 95.2%; Pred. NO. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21
    |||||||||||||||
Db 664 GCTTGGACTATAATACCTGAC 684

RESULT 8
LOCUS      ATTDNA 878 bp DNA
DEFINITION Agrobacterium tumefaciens crown gall tumor T-DNA from Ti (tumor
            inducing) plasmid pTiA6.
ACCESSION  X00431
VERSION    X00431.1 GI:39150
KEYWORDS   plasmid.
SOURCE     Agrobacterium tumefaciens.
ORGANISM   Agrobacterium tumefaciens.
            Bacteria; Proteobacteria;
            Rhizobiaceae; Rhizobium.
            1 (bases 1 to 878)
            McPherson,J.C.
            DNA sequence analysis of crown gall tumor T-DNA encoding the 0.7 kb
            transcript
            Nucleic Acids Res. 12 (5), 2317-2325 (1984)
JOURNAL    84169535
MEDLINE
```

FEATURES  
source Location/Qualifiers  
1..878  
/organism="Agrobacterium tumefaciens"  
/strain="plasmid pRIa6"  
/db\_xref="taxon:358"  
60..66  
/note="TATA-box"  
68..613  
/note="polyadenylation signal"  
109..489  
/note="unidentified reading frame"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA25129.1"  
/db\_xref="GI:39151"  
/db\_xref="SWISS-PROT:P03867"  
/translation="MNEADTFLASLDLMDACEEFIKTYGASPOLETGEVIQTNNGLLY  
LYGKGSLSQRHDTFLAFKEKEELSFTTIKPAEMKAQOSDLTYVVAIFQSNFLCVSN  
PERGLRCHNRPFLYPIVAGHSM5"  
659..664  
/note="polyadenylation signal"  
BASE COUNT 288 a 189 c 139 g 262 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 1; Length 878;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 625 GCTTGGACTATAACTCTGAC 645

RESULT 9  
AX172475/c  
LOCUS AX172475 909 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 36 from Patent WO0141558.  
ACCESSION AX172475  
VERSION AX172475.1 GI:14597587  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS de Both, G. and de Beuckeleer, M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 36 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source Location/Qualifiers  
1..909  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="sequence comprising the 5' flanking region of MS-BN1 in WOSR"

BASE COUNT 309 a 134 c 173 g 293 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 909;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 909 GCTTGGACTATAACTCTGAC 889

RESULT 10  
AX10942  
LOCUS AX10942 1037 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 4 from patent number DE3920034.  
ACCESSION AX10942

VERSION A10942.1 GI:492369  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1037)  
AUTHORS  
JOURNAL Patent: DE 3920034-A 4 31-MAY-1990;  
FEATURES Location/Qualifiers  
1..1037  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 338 a 174 c 166 g 359 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1037;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 921 GCTTGGACTATAACTCTGAC 941

RESULT 11  
AX172463/c  
LOCUS AX172463 1077 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 24 from Patent WO0141558.  
ACCESSION AX172463  
VERSION AX172463.1 GI:14597575  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 1077)  
AUTHORS de Both, G. and de Beuckeleer, M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 24 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source Location/Qualifiers  
1..1077  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="sequence comprising the 5' flanking region of RF-BN1"

misc\_feature 1..45  
misc\_feature 1061..1077  
BASE COUNT 430 a 148 c 151 g 347 t  
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1077;  
Best Local Similarity 95.2%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 980 GCTTGGACTATAACTCTGAC 960

RESULT 12  
A10939  
LOCUS A10939 1085 bp DNA PAT 27-SEP-1993  
DEFINITION Nucleotide sequence 1 from patent number DE3920034.  
ACCESSION A10939  
VERSION A10939.1 GI:492367  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.

```

REFERENCE 1 (bases 1 to 1085)
AUTHORS
JOURNAL
FEATURES
  source
    1..1085
      /organism="unidentified"
      /db_xref="taxon:32644"
BASE COUNT 369 a 218 c 155 g 343 t
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1085;
Best Local Similarity 95.2%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataatacttgac 21
|||||
Db 969 GCTTGGACTATAATACCTGAC 989

RESULT 13
AL0943
LOCUS AL0943 1160 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 5 from patent number DE3920034.
ACCESSION AL0943
VERSION AL0943.1 GI:492370
KEYWORDS
SOURCE
  unidentified.
  unclassified.
ORGANISM
  unclassified.
REFERENCE 1 (bases 1 to 1160)
AUTHORS
JOURNAL
FEATURES
  source
    1..1160
      /organism="unidentified"
      /db_xref="taxon:32644"
BASE COUNT 367 a 194 c 188 g 411 t
ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1160;
Best Local Similarity 95.2%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataatacttgac 21
|||||
Db 1044 GCTTGGACTATAATACCTGAC 1064

RESULT 14
AL18051
LOCUS AL18051 1186 bp DNA PAT 26-JUL-1994
DEFINITION DNA used as a probe for neo gene seq ID No:2.
ACCESSION AL18051
VERSION AL18051.1 GI:583120
KEYWORDS
  synthetic construct.
  synthetic construct.
  artificial sequence.
SOURCE
  1 (bases 1 to 1186)
ORGANISM
  1..1186
    Location/Qualifiers
      /organism="synthetic construct"
      /db_xref="taxon:32630"
REFERENCE 1 (bases 1 to 1186)
AUTHORS
JOURNAL
FEATURES
  source
    WO 9209696-A 2 11-JUN-1992;
    Location/Qualifiers
      1..1186
        /organism="synthetic construct"
        /db_xref="taxon:32630"
  promoter
    1..8
      /note="sequence derived from tapetum specific promoter of
      Nicotiana tabacum"
  gene
    167..790
      /gene="neomycine phosphotransferase gene"
  CDS
    167..790

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:40 ; Search time 716.55 Seconds  
(without alignments)  
25.126 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataataacttgac 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	21	100.0	21	AAD07001	PCR primer B02, to
2	19.4	92.4	21	AAH25431	PCR primer for vec
3	19.4	92.4	415	AAD06997	Right (5') border
4	19.4	92.4	416	AAD06999	Left (3') border f
5	19.4	92.4	694	AAH25454	Nucleotide fragmen
6	19.4	92.4	909	AAH25451	Nucleotide fragmen
7	19.4	92.4	1037	AAQ04705	USP-Promoter-casse
8	19.4	92.4	1077	AAH25439	Right flanking reg
9	19.4	92.4	1085	AAQ04703	Legumin-signalpept
10	19.4	92.4	1160	AAQ04706	USP-signalpeptide
11	19.4	92.4	1186	AAQ25707	Chimeric neo gene

12	19.4	92.4	1303	17	AAT39337	Plasmid pT888 (Eco
13	19.4	92.4	3153	21	AAZ29122	plasmid DV131 comp
14	19.4	92.4	3201	12	AAQ14529	ppS029 Bt ICP codi
15	19.4	92.4	3201	12	AAQ15144	pVE36 Bt ICP codin
16	19.4	92.4	3336	21	AAZ29121	Plasmid DV130 comp
17	19.4	92.4	3694	21	AAZ29124	Plasmid DV133 used
18	19.4	92.4	3877	21	AAZ29123	Plasmid DV132 used
19	19.4	92.4	4832	22	AAH25423	Nucleotide sequenc
20	19.4	92.4	4946	18	AAT59531	T-DNA of plasmid p
21	19.4	92.4	4946	18	AAT59531	Nucleotide sequenc
22	19.4	92.4	5228	22	AAH25432	Plasmid pT5172delc
23	19.4	92.4	5349	19	AAZ23239	T-DNA of pT524.
24	19.4	92.4	5864	17	AAT39339	Plasmid pT0113 T-
25	19.4	92.4	5864	17	AAT39339	Plasmid pT0113 T-
26	19.4	92.4	5865	22	AAD06990	Chimeric T-DNA of
27	19.4	92.4	5865	22	AAD06990	Chimeric T-DNA of
28	19.4	92.4	6539	21	AAZ91097	E. coli plasmid pT
29	19.4	92.4	6548	17	AAT39336	Plasmid pT5174 use
30	19.4	92.4	6548	18	AAT61394	Plasmid pT5172. C
31	19.4	92.4	6548	21	AAZ91096	E. coli plasmid pT
32	19.4	92.4	7492	22	AAH86441	Plasmid pT3346. U
33	19.4	92.4	7565	14	AAQ42160	Plasmid pPS0212 co
34	19.4	92.4	7599	22	AAH25320	Nucleotide sequenc
35	19.4	92.4	7639	14	AAQ42159	Plasmid pJ0884 con
36	19.4	92.4	24593	6	AAH50226	Sequence of opine
37	19.4	92.4	24596	6	AAH50182	Complete nucleotid
38	16.2	77.1	423	21	AAAC02009	Human secreted pro
39	16	76.2	22	20	AAZ28158	Transformed Arabid
40	15.8	75.2	81	16	AAH06276	HIV-1 reverse tran
41	15.8	75.2	359	21	AAH06027	Human secreted pro
42	15.8	75.2	831	20	AAZ10777	Trehalose-6-phosph
43	15.8	75.2	909	20	AAZ22113	Human secreted pro
44	15.8	75.2	1433	15	AAQ65397	Elm C10:O-acyl car
45	15.8	75.2	1433	16	AAQ92306	Elm class II thioe

ALIGNMENTS

RESULT 1

AAD07001 AAD07001 standard; DNA; 21 BP.  
XX AC AAD07001;  
XX DT 06-AUG-2001 (first entry)  
XX DE PCR primer B02, to recognise foreign DNA and flanking sequence of MS-B2.  
XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
XX KW male-sterility gene; PCR primer; ss.  
XX OS Unidentified.  
XX PN WO200131042-A2.  
XX PD 03-MAY-2001.  
XX PF 26-OCT-2000; 2000WO-EP10680.  
XX PR 29-OCT-1999; 99US-0430497.  
XX PA (AVET ) AVENTIS CROPS SCIENCE NV.  
XX PI Weston B, De Beuckeleer M;  
XX DR WPI; 2001-300517/31.  
XX PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
XX PT harboring specific transformation events, particularly by presence of  
XX PT male-sterility gene, at specific location in its genome -  
XX PS Claim 1; Page 33; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.  
 XX Sequence 21 BP; 6 A; 4 C; 4 G; 7 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0;  
 QY 1 gcttggactataacttgac 21  
 |||||  
 Db 1 gcttggactataacttgac 21

RESULT 2  
 AAH25431  
 ID AAH25431 standard; DNA; 21 BP.  
 AC AAH25431;  
 XX  
 DT 22-AUG-2001 (first entry)  
 DE PCR primer for vector fragments in transgenic plant MS-BN1.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.  
 OS Synthetic.  
 XX  
 PN WO200141558-A1.  
 XX  
 PD 14-JUN-2001.

XX 06-DEC-2000; 2000WO-EP12872.  
 XX  
 PF 08-DEC-1999; 99US-0457037.  
 XX  
 PR (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA  
 PI De Both G, De Beuckeleer M;  
 XX WPI; 2001-381419/40.  
 DR  
 XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome -  
 XX  
 PS Example 4; Page 43; 98pp; English.

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH25429-31 were used to amplify fragments of a vector in  
 CC a transgenic plant which carries the TA29-barnase transgene.  
 XX Sequence 21 BP; 6 A; 5 C; 4 G; 6 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 21;  
 Best Local Similarity 95.2%; Pred. No. 1.1; Mismatches 1; Indels 0; Gaps 0;  
 Matches 20; Conservative 0;  
 QY 1 gcttggactataacttgac 21  
 |||||  
 Db 1 gcttggactataacttgac 21

RESULT 3  
 AAD06997/c  
 ID AAD06997 standard; DNA; 415 BP.  
 XX  
 AC AAD06997;  
 XX  
 DT 06-AUG-2001 (first entry)  
 XX  
 DE Right (5') border flanking region of elite event MS-B2.  
 KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; ds.  
 XX Chimeric - Agrobacterium sp.  
 OS Chimeric - Brassica sp.

XX Key Location/Qualifiers  
 FT misc\_feature 1..234  
 FT /\*tag= a  
 FT /note= "Corresponds to plant DNA"  
 FT 235..415  
 FT /\*tag= b  
 FT /note= "Corresponds to T-DNA"

XX WO200131042-A2.  
 PN  
 XX 03-MAY-2001.  
 PD  
 XX 26-OCT-2000; 2000WO-EP10680.  
 PF  
 XX 29-OCT-1999; 99US-0430497.  
 PR  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA  
 XX Weston B, De Beuckeleer M;  
 PI WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harbouring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX  
 PS Claim 11; Page 51; 53pp; English.  
 XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is right (5') border flanking region of elite event  
 CC MS-B2.  
 XX Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 415;  
 Best Local Similarity 95.2%; Pred. No. 1.3; Mismatches 1; Indels 0; Gaps 0;  
 Matches 20; Conservative 0;  
 QY 1 gcttggactataacttgac 21  
 |||||  
 Db 335 GCTTGGACTATAACTTGAC 315



```

RESULT 4
AAD06999
ID AAD06999 standard; DNA; 416 BP.
XX
AC AAD06999;
XX
DT 06-AUG-2001 (first entry)
XX
DE Left (3') border flanking region of elite event MS-B2.
XX
KW MS-B2 elite event; transgenic Brassica plant; transformation event;
KW male-sterility gene; ds.
XX
OS Chimeric - Agrobacterium sp.
OS Chimeric - Brassica sp.
XX
FH Key Location/Qualifiers
FT misc_feature 1..193
FT /*tag= a
FT /note= "Corresponds to T-DNA"
FT misc_feature 194..416
FT /*tag= b
FT /note= "Corresponds to plant DNA"
XX
PN WO200131042-A2.
XX
PD 03-MAY-2001.
XX
PF 26-OCT-2000; 2000WO-EP10680.
XX
PR 29-OCT-1999; 99US-0430497.
XX
PA (AVET ) AVENTIS CROPS SCIENCE NV.
XX
PI Weston B, De Beuckeleer M;
DR WPI; 2001-300517/31.
XX
Transgenic Brassica plants, seeds, cells or tissues, characterized by
PT harboring specific transformation events, particularly by presence of
PT male-sterility gene, at specific location in its genome -
XX
Claim 11; Page 52; 53pp; English.
XX
The present invention relates to a transgenic Brassica plant or its
CC seed, cells or tissues, characterised by harbouring a specific
CC transformation event, particularly by the presence of a male-sterility
CC gene, at a specific location in the Brassica genome. Transgenic
CC Brassica plant is useful for producing a hybrid seed by crossing the
CC transgenic plant with a male-fertile Brassica plant and harvesting the
CC hybrid seed from the transgenic Brassica plant.
CC The present sequence is left (3') border flanking region of elite event
CC MS-B2.
XX
SQ Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;

Query Match 92.4%; Score 19.4; DB 22; Length 416;
Best Local Similarity 95.2%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21
Db 81 gcttgactataactcgac 101

RESULT 5
AAH25454/C
ID AAH25454 standard; DNA; 694 BP.
XX
AC AAH25454;

Query Match 92.4%; Score 19.4; DB 22; Length 416;
Best Local Similarity 95.2%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21
Db 81 gcttgactataactcgac 101

RESULT 6
AAH25451/C
ID AAH25451 standard; DNA; 909 BP.
XX
AC AAH25451;
XX
DT 22-AUG-2001 (first entry)
XX
DE Nucleotide fragment of a vector comprising TA29-barnase transgene.
XX
Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barnase gene; barstar gene; ss.
XX
OS Synthetic.
XX
PN WO200141558-A1.
XX
PD 14-JUN-2001.
XX
SQ Sequence 694 BP; 298 A; 81 C; 73 G; 242 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 694;
Best Local Similarity 95.2%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21
Db 694 GCTTGACTATATACCTGAC 674

RESULT 7
AAH25451/C
ID AAH25451 standard; DNA; 909 BP.
XX
AC AAH25451;
XX
DT 22-AUG-2001 (first entry)
XX
DE Nucleotide fragment of a vector comprising TA29-barnase transgene.
XX
Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barnase gene; barstar gene; ss.
XX
OS Synthetic.
XX
PN WO200141558-A1.
XX
PD 14-JUN-2001.
XX
SQ Sequence 694 BP; 298 A; 81 C; 73 G; 242 T; 0 other;

The specification describes a pair of transgenic winter oilseed rape
plants suited for producing hybrid seed. One of the plants has an
expression cassette comprising a male-sterility gene (e.g. barnase
gene), and the other plant has an expression cassette comprising a
fertility restorer gene (e.g. barstar gene), integrated into the genome.
The fertility restorer gene is capable of preventing the activity of the
male-sterility gene. The plant pair is useful for producing hybrid seed.
Plants developed from the hybrid seed have agronomic performance,
genetic stability and adaptability to different genetic backgrounds.
The present sequence represents a fragment from a vector comprising the
TA29-barnase transgene from transgenic plants.

Sequence 694 BP; 298 A; 81 C; 73 G; 242 T; 0 other;

```

PF 06-DEC-2000; 2000WO-EP12872.  
 XX  
 PR 08-DEC-1999; 99US-0457037.  
 XX  
 PA (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX  
 XX De Both G, De Beuckeleer M;  
 XX WPI; 2001-381419/40.  
 DR  
 XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome  
 XX  
 XX Example 4; Page 93; 98pp; English.  
 PS  
 XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present sequence represents a fragment from a vector comprising the  
 CC TA29-barnase transgene from transgenic plants.  
 XX  
 XX Sequence 909 BP; 309 A; 134 C; 173 G; 293 T; 0 other;

Query Match 92.4%; Score 19.4; DB 22; Length 909;  
 Best Local Similarity 95.2%; Pred. No. 1.3;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgagctataataactgtac 21  
 |||||  
 Db 909 GCTTGACTATAATACCTGAC 889

RESULT 7  
 AAQ04705  
 ID AAQ04705 standard; DNA; 1037 BP.  
 XX  
 AC AAQ04705;  
 XX  
 DT 12-OCT-1990 (first entry)  
 XX  
 DE USP-Promoter-cassette USP-Pr.T7.1.  
 XX  
 KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.  
 XX  
 PN DE3920034-A.  
 XX  
 PD 31-MAY-1990.  
 XX  
 PF 20-JUN-1989; 89DE-3920034.  
 XX  
 XX 19-SEP-1988; 88DD-0319887.  
 XX (PFLA-) VE KOMB PFLANZENZUC.  
 XX  
 XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 XX WPI; 1990-172459/23.  
 DR  
 XX Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 PT  
 XX Disclosure; ; pp; German.  
 PS  
 XX The unique BglII-Ort (720-725) site is for ligating foreign DNA and the

CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the T1-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 XX Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 92.4%; Score 19.4; DB 11; Length 1037;  
 Best Local Similarity 95.2%; Pred. No. 1.4;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgagctataataactgtac 21  
 |||||  
 Db 921 gcttgagctataataactgtac 941

RESULT 8  
 AAH25439/C  
 ID AAH25439 standard; DNA; 1077 BP.  
 XX  
 AC AAH25439;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Right flanking region in transgenic plant BN-RFL.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; barstar gene; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200141558-A1.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 06-DEC-2000; 2000WO-EP12872.  
 XX  
 XX 08-DEC-1999; 99US-0457037.  
 PR  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA  
 XX De Both G, De Beuckeleer M;  
 PI  
 XX WPI; 2001-381419/40.  
 DR  
 XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
 PT with improved qualities, comprises a male-sterility gene and fertility  
 PT restorer gene, integrated into the genome  
 XX  
 XX Claim 58; Page 88-89; 98pp; English.  
 PS  
 XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present sequence represents the right flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barstar  
 CC transgene.  
 XX  
 XX Sequence 1077 BP; 430 A; 148 C; 151 G; 347 T; 1 other;

Query Match 92.4%; Score 19.4; DB 22; Length 1077;  
 Best Local Similarity 95.2%; Pred. No. 1.4;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 gcttgagctataataactgtac 21

Db 980 GCTTGGACTATAATACCTGAC 960  
|||||

## RESULT 9

AAQ04703  
ID AAQ04703 standard; DNA; 1085 BP.

XX AC AAQ04703;

XX DT 12-OCT-1990 (first entry)

XX DE Legumin-signalpeptide cassette Le-Sig.T7.

XX KW Foreign DNA incorporation; recombinant DNA techniques;  
XX KW higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.

XX FH Key Location/Qualifiers

XX CDS 747..814

XX FT /\*tag= a  
XX FT /product=Legumin-signalpeptide

XX PN DE3920034-A.

XX PD 31-MAY-1990.

XX PF 20-JUN-1989; 89DE-3920034.

XX PR 19-SEP-1988; 88DD-0319887.

XX PA (PFLA-) VE KOMB PFLANZENZUC.

XX PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

XX DR WPI; 1990-172459/23.

XX DR P-PSDB; AAR05198.

XX PT Incorporation of DNA into higher plant genome - by specified  
XX PT recombinant DNA techniques.

XX PS Disclosure; ; pp; German.

XX CC The Legumin gene B4 is used. The unique BglII-Ort (815-820) site is for  
XX CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
XX CC (1080-1085) for cloning the cassette in the Ti-vector pGA471.  
XX CC The cassette is cloned into the binary Ti-vectors pGA471 and  
XX CC Agrobacterium tumefaciens is transfected.  
XX CC See also AAQ04703-Q04706.

XX SQ Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;

## Query Match

Best Local Similarity 92.4%; Score 19.4; DB 11; Length 1085;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21

Db 969 gcttggactataactcgac 989

## RESULT 10

AAQ04706

ID AAQ04706 standard; DNA; 1160 BP.

XX AC AAQ04706;

XX DT 12-OCT-1990 (first entry)

XX DE USP-signalpeptide cassette USP-Sig.T7.

XX KW Foreign DNA incorporation; recombinant DNA techniques;  
XX KW higher plant genome; signalpeptide; USP-Sig.T7.; ss.

XX Key Location/Qualifiers  
XX CDS 708..877  
XX FT /\*tag= a  
XX FT /product=signalpeptide  
XX FT 747..817

XX PN intron

XX PN DE3920034-A.

XX PD 31-MAY-1990.

XX PF 20-JUN-1989; 89DE-3920034.

XX PR 19-SEP-1988; 88DD-0319887.

XX PA (PFLA-) VE KOMB PFLANZENZUC.

XX PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

XX DR WPI; 1990-172459/23.

XX DR P-PSDB; AAR05199.

XX PT Incorporation of DNA into higher plant genome - by specified  
XX PT recombinant DNA techniques.

XX PS Disclosure; ; pp; German.

XX CC The unique BglII-Ort (890-895) site is for  
XX CC ligating foreign DNA and the HindIII-Ort in the 3' polylinker  
XX CC (1155-1160) for cloning the cassette in the Ti-vector pGA471.  
XX CC The cassette is cloned into the binary Ti-vectors pGA471 and  
XX CC Agrobacterium tumefaciens is transfected.  
XX CC See also AAQ04703-Q04706.

XX SQ Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;

## Query Match

Best Local Similarity 92.4%; Score 19.4; DB 11; Length 1160;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataacttgac 21

Db 1044 gcttggactataactcgac 1064

## RESULT 11

AAQ25707

ID AAQ25707 standard; DNA; 1186 BP.

XX AC AAQ25707;

XX DT 07-DEC-1992 (first entry)

XX DE Chimeric neo gene probe.

XX KW Nicotiana tabacum; neomycine phosphotransferase; Agrobacterium; ss.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT promoter 1..8

XX FT /\*tag= a  
XX FT /note= "sequence derived from tapetum specific  
promoter of Nicotiana tabacum"

XX FT CDS 9..790

XX FT /\*tag= b  
XX FT /product= neomycine\_phosphotransferase

XX FT misc\_feature 791..1186

XX FT /\*tag= c

XX FT /note= "3' regulatory sequence contg. the  
polyadenylation site derived from  
Agrobacterium T-DNA gene 7"

```

XX PN W09209696-A.
XX PD 11-JUN-1992.
XX PF 21-NOV-1991; 91WO-EP02198.
XX PR 23-NOV-1990; 90EP-0403332.
XX PR 08-JUL-1991; 91EP-0401888.
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PI Dhalluin K, Goebel E;
XX DR WPI; 1992-217075/26.
XX XX Transforming monocotyledonous plants e.g. cereals - comprises
XX PT wounding and/or degrading cells of intact plant tissue or
XX PT embryogenic callus
XX PS Disclosure; Page 60; 76pp; English.
XX CC Two transformed corn plants were analysed by means of Southern
XX CC hybridisation. As a probe, a 1184 bp EcoRI-HindIII fragment derived
XX CC from another plasmid was used. The sequence of that plasmid is
XX CC given below. Results showed that at least a chimeric neo gene was
XX CC integrated into the plant genomic DNA.
XX SQ Sequence 1186 BP; 244 A; 317 C; 325 G; 300 T; 0 other;

Query Match 92.4%; Score 19.4; DB 13; Length 1186;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 gcttgactataacttgac 21
   |||||
DB 1113 gcttgactataacttgac 1133

RESULT 12
AAT39337
ID AAT39337 standard; DNA; 1303 BP.
AC AAT39337;
XX
XX 22-JAN-1997 (first entry)
XX Plasmid pTS88 (EcoRI-HindIII fragment).
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
XX transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX misc_feature 1..35
   /tag= a
   /label= pGEM2
   /note= "polylinker of pGEM2"
   /tag= b
   /label= p35S
   /function= 35S promoter of cauliflower mosaic virus
   strain CM1841
CDS 695..967
   /tag= c
   /label= barstar
   /product= Bacillus amyloliquefaciens barstar
FT polyA_signal 968..1287
   /tag= d
   /label= 3'g7
   /function= region containing polyadenylation signal

```

```

FT misc_feature 1288..1303
FT /tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
XX W09626283-A1.
XX PD 29-AUG-1996.
XX PF 21-FEB-1996; 96WO-EP00722.
XX PR 21-FEB-1995; 95EP-0400364.
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PI Botterman J, Cornelissen M, Michiels F;
XX DR WPI; 1996-402373/40.
XX XX Prodn. of male sterile plants by transforming with a chimaeric
XX PT construct - comprising a male sterility DNA e.g. barnase and a
XX PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX PT for generating hybrid cultivars
XX PS Example 1; Page 38; 56pp; English.
XX CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
XX CC barstar DNA under control of a 35S promoter. The plasmid was
XX CC used with pTS174 (see also AAT39336) contg. barnase DNA under
XX CC control of the stamen-specific promoter E1 to produce male sterile
XX CC rice cv. Kochinibiki transgenic plants, and with plasmid pV136
XX CC (see also AAT39338) contg. barnase DNA under control of the stamen-
XX CC specific PCA55 promoter to produce male sterile maize plants.
XX CC Expression of barnase (a ribonuclease) in the stamen leads to male
XX CC sterility. Constitutive expression of barstar counteracts possible
XX CC low level expression of barnase DNA in non-stamen tissue.
XX SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 92.4%; Score 19.4; DB 17; Length 1303;
Best Local Similarity 95.2%; Pred. No. 1.4;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 gcttgactataacttgac 21
   |||||
DB 1178 gcttgactataacttgac 1198

RESULT 13
AAZ29122
ID AAZ29122 standard; DNA; 3153 BP.
XX
XX AAZ29122;
XX AC AAZ29122;
XX DT 21-FEB-2000 (first entry)
XX DE Plasmid DV131 comprising L3/EGFP:NPTII/Tr7 expression cassette.
XX KW Transgenic seed; marker; aleurone-specific promoter; plasmid DV131;
XX KW GFP:NPTII fusion protein construct; L3/EGFP:NPTII/Tr7; maize R gene;
XX KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;
XX KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene; assay;
XX KW Green fluorescent protein; GFP; gene fusion; selection; screening;
XX KW expression; automated seed screening technique; screenable marker;
XX KW transformant; embryogenic tissue; implementation; ds.
XX OS Synthetic.
XX XX
XX PN W0960129-A1.
XX PD 25-NOV-1999.

```

XX 18-MAY-1999; 99WO-US11023.  
 XX  
 XX 18-MAY-1998; 98US-0080625.  
 XX  
 XX (DEKA-) DEKALB GENETICS CORP.  
 XX  
 XX Kriz AL, Spencer TM;  
 XX  
 XX WPI; 2000-072441/06.  
 XX

PT Screenable marker genes useful for identification of transgenic seeds  
 for plant breeding -

XX Example 1; Page 164-166; 182pp; English.

XX The present DNA sequence is the plasmid DVI131, that is used in the  
 CC generation of GFP:NPTII fusion protein constructs. It contains an  
 CC expression cassette comprising, a promoter from the maize L3 oleosin  
 CC gene, the coding sequence of EGFP:NPTII translational fusion, excised  
 CC from DVI126 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to the gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.

XX Sequence 3153 BP; 801 A; 857 C; 756 G; 739 T; 0 other;

Query Match 92.4%; Score 19.4; DB 21; Length 3153;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataataacttgac 21  
 |||||

Db 2759 gcttggactataataacttgac 2779

RESULT 14  
 AAQ14529  
 ID AAQ14529 standard; DNA; 3201 BP.

XX AAQ14529;

XX 27-JAN-1992 (first entry)

XX pPS029 Bt ICP coding sequence.

XX Bacillus thuringiensis; insecticidal crystal protein; ICP;  
 KW deletion; ss.

XX Synthetic.

XX WO9116432-A.

XX 31-OCT-1991.

XX 17-APR-1991; 91WO-EP00733.

XX 18-APR-1990; 90EP-0401055.

XX (PLAN-) PLANT GENETIC SYST.

XX Cornelissen M, Soetaert P, Stam M, Dockx J;

XX WPI; 1991-339820/46.

XX

PT Modified Bacillus thuringiensis insecticidal crystal protein  
 PT genes - having A and T sequences changed to G and C sequences  
 XX encoding same amino acids, for increased expression levels  
 XX Disclosure; Fig 6(c); 78pp; English.

XX "n" in the sequence refers to not known nucleotides.  
 CC pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-  
 CC terminal modification and the internal modification of the Bt ICP  
 CC coding sequence.  
 CC See also AAQ14529, AAQ15142-44.

XX Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;

Query Match 92.4%; Score 19.4; DB 12; Length 3201;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataataacttgac 21  
 |||||

Db 3032 gcttggactataataacttgac 3052

RESULT 15

AAQ15144

ID AAQ15144 standard; DNA; 3201 BP.

XX AAQ15144;

XX 27-JAN-1992 (first entry)

XX pVE36 Bt ICP coding sequence.

XX Bacillus thuringiensis; insecticidal crystal protein; ICP;  
 KW deletion; ss.

XX Synthetic.

XX WO9116432-A.

XX 31-OCT-1991.

XX 17-APR-1991; 91WO-EP00733.

XX 18-APR-1990; 90EP-0401055.

XX (PLAN-) PLANT GENETIC SYST.

XX Cornelissen M, Soetaert P, Stam M, Dockx J;

XX WPI; 1991-339820/46.

XX Modified Bacillus thuringiensis insecticidal crystal protein  
 XX genes - having A and T sequences changed to G and C sequences  
 XX encoding same amino acids, for increased expression levels

XX Disclosure; Fig 6(c); 78pp; English.

XX "n" in the sequence refers to not known nucleotides.

XX pPS029 (AAQ14529) is identical to pVE36, but carries both the amino-  
 XX terminal modification and the internal modification of the Bt ICP  
 XX coding sequence.

XX See also AAQ14529, AAQ15142-44.

XX Sequence 3201 BP; 915 A; 651 C; 700 G; 930 T; 5 other;

Query Match 92.4%; Score 19.4; DB 12; Length 3201;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataataacttgac 21

Tue Feb 26 09:11:31 2002

us-09-698-903b-12.rng

Page 8

Db 3105 gcttgactatatacctgac 3125  
|||||

Search completed: February 25, 2002, 18:17:41  
Job time: 16699 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:51 ; Search time 301.6 Seconds  
(without alignments)  
15.769 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataacttgac 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/1na/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/1na/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	92.4	1186	1	US-08-064-121-2
2	19.4	92.4	1186	1	US-08-478-015-2
3	19.4	92.4	1186	3	US-08-475-975-2
4	19.4	92.4	1186	3	US-09-084-889-2
5	19.4	92.4	1303	3	US-08-894-440-2
6	19.4	92.4	1353	4	US-09-080-625-3
7	19.4	92.4	3200	1	US-08-453-104-23
8	19.4	92.4	3200	2	US-08-694-824-23
9	19.4	92.4	3201	1	US-08-453-104-22
10	19.4	92.4	3201	2	US-08-694-824-22
11	19.4	92.4	3336	4	US-09-080-625-2
12	19.4	92.4	3694	4	US-09-080-625-5
13	19.4	92.4	3877	4	US-09-080-625-4
14	19.4	92.4	4946	3	US-08-817-188-1
15	19.4	92.4	5560	3	US-08-817-188-5
16	19.4	92.4	5864	3	US-08-894-440-4
17	19.4	92.4	5864	3	US-08-894-440-4
18	19.4	92.4	6548	3	US-08-894-440-1
19	19.4	92.4	6548	3	US-08-817-188-2
20	19.4	92.4	7566	2	US-08-232-016-23
21	19.4	92.4	7639	2	US-08-232-016-22
22	19.4	92.4	7811	2	US-08-549-680A-5
23	19.4	92.4	24595	6	5428147-1
24	15.8	75.2	81	1	US-08-238-863-88
25	15.8	75.2	81	1	US-08-443-407-88
26	15.8	75.2	81	5	PCT-US95-05600-232
27	15.8	75.2	1433	1	US-07-968-971A-11

c 28	15.8	75.2	1433	1	US-08-383-756-5	Sequence 5, Appl
c 29	15.8	75.2	1433	1	US-08-424-406-2	Sequence 2, Appl
c 30	15.8	75.2	1433	1	US-08-464-523B-8	Sequence 8, Appl
c 31	15.8	75.2	1433	2	US-08-460-898-5	Sequence 5, Appl
c 32	15.2	72.4	262	4	US-09-437-457-13	Sequence 13, Appl
c 33	15.2	72.4	1618	1	US-08-410-540-1	Sequence 1, Appl
c 34	15.2	72.4	1641	4	US-08-659-254-19	Sequence 19, Appl
c 35	15.2	72.4	2811	4	US-08-482-073-3	Sequence 3, Appl
c 36	15.2	72.4	2813	2	US-08-344-155C-99	Sequence 99, Appl
c 37	15.2	72.4	2813	4	US-09-009-490A-90	Sequence 90, Appl
c 38	15.2	72.4	3080	4	US-08-482-073-4	Sequence 4, Appl
c 39	15.2	72.4	4016	1	US-08-410-540-3	Sequence 3, Appl
c 40	14.8	70.5	1925	4	US-08-894-324A-1	Sequence 1, Appl
c 41	14.8	70.5	2778	1	US-08-202-054-1	Sequence 1, Appl
c 42	14.8	70.5	2778	1	US-08-446-923-1	Sequence 1, Appl
c 43	14.6	69.5	3390	4	US-09-550-338-1	Sequence 1, Appl
c 44	14.4	68.6	1498	1	US-07-965-688A-1	Sequence 1, Appl
c 45	14.4	68.6	1498	2	US-08-950-433-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-064-121-2  
; Sequence 2, Application US/08064121  
; Patent No. 5641664  
; GENERAL INFORMATION:  
; APPLICANT: D'HALLUIN, Kathleen  
; APPLICANT: GOBEL, Elke  
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/064,121  
; FILING DATE: 24-MAY-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 90403332.1  
; FILING DATE: 23-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91401888.2  
; FILING DATE: 08-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crane-Feuiv, Sharon E  
; REGISTRATION NUMBER: 36,113  
; REFERENCE/DOCKET NUMBER: 010830-043  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1186 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORGANISM: probe

```

; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
US-08-064-121-2

```

```

Query Match          92.4%; Score 19.4; DB 1; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 gcttgactataactctgac 21
|||||
Db 1113 GCTTGACTATAACTCTGAC 1133

```

```

RESULT 2
US-08-478-015-2
; Sequence 2, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P. O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine phosphotransferase ge
US-08-478-015-2

```

```

Query Match          92.4%; Score 19.4; DB 1; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 gcttgactataactctgac 21
|||||
Db 1113 GCTTGACTATAACTCTGAC 1133

```

```

RESULT 3
US-08-475-975-2
; Sequence 2, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:

```



```
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: monocytyledonous plants
; OTHER INFORMATION: T-DNA gene 7"
; US-08-475-975-2

Query Match 92.4%; Score 19.4; DB 3; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataactctgac 21
|||||
Db 1113 GCTTGGACTATAACTCTGAC 1133

RESULT 4
US-09-084-889-2
; Sequence 2, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/084,889
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; FILING DATE:
; PRIOR APPLICATION NUMBER: 08/064,121
; APPLICATION DATA: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-09-084-889-2

Query Match 92.4%; Score 19.4; DB 3; Length 1186;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttggactataactctgac 21
|||||
Db 1113 GCTTGGACTATAACTCTGAC 1133

RESULT 5
US-08-894-440-2
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2:
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(35)
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36)..(694)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (695)..(967)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amy1oliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (968)..(1287)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1288)..(1303)
; OTHER INFORMATION: polylinker of pGEM2
; US-08-894-440-2
```

```
Query Match          92.4%; Score 19.4; DB 3; Length 1303;
Best Local Similarity 95.2%; Pred. No. 0.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 gcttgactataactgtac 21
      |||||
Db 1178 gcttgactataactgtac 1198
```

```
RESULT 6
US-09-080-625-3
; Sequence 3, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/080,625
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
; US-09-080-625-3
;
; Query Match          92.4%; Score 19.4; DB 4; Length 3153;
; Best Local Similarity 95.2%; Pred. No. 0.32;
; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 gcttgactataactgtac 21
;      |||||
; Db 2759 GCTTGGACTATAACTGTAC 2779
;
; RESULT 7
; US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/453,104
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
; US-08-453-104-23
```

```
Query Match          92.4%; Score 19.4; DB 1; Length 3200;
Best Local Similarity 95.2%; Pred. No. 0.32;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 gcttgactataacttgac 21  
|||||  
Db 3032 GCTTGACTATAACTCGAC 3052

## RESULT 8

US-08-694-824-23  
; Sequence 23, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 92.4%; Score 19.4; DB 2; Length 3200;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 3032 GCTTGACTATAACTCGAC 3052

## RESULT 9

US-08-453-104-22  
; Sequence 22, Application US/08453104

; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 92.4%; Score 19.4; DB 1; Length 3201;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataacttgac 21  
|||||  
Db 3105 GCTTGACTATAACTCGAC 3125

## RESULT 10

US-08-694-824-22  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

;; TITLE OF INVENTION: IN PLANT CELLS  
;; NUMBER OF SEQUENCES: 23  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Burns, Doane, Swecker & Mathis  
;; STREET: George Mason Bldg., Washington & Prince Sts.  
;; CITY: Alexandria  
;; STATE: Virginia  
;; COUNTRY: United States  
;; ZIP: 22313-1404  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/694,824  
;; FILING DATE: 09-AUG-1996  
;; CLASSIFICATION: 536  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/937,869  
;; FILING DATE: 16-DEC-1992  
;; APPLICATION NUMBER: GB 90401055.0  
;; FILING DATE: 18-APR-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Rea, Teresa S  
;; REGISTRATION NUMBER: 30,427  
;; REFERENCE/DOCKET NUMBER: 010830-032  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703) 836-6620  
;; TELEFAX: (703) 836-2021  
;; INFORMATION FOR SEQ ID NO: 22:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3201 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 2151..2155  
;; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
;; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 92.4%; Score 19.4; DB 2; Length 3201;  
Best Local Similarity 95.2%; Pred. No. 0.32;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||  
Db 3105 GCTTGGACTATAACTGTGAC 3125

RESULT 11  
US-09-080-625-2  
; Sequence 2, Application US/09080625  
; Patent No. 6307123  
; GENERAL INFORMATION:  
; APPLICANT: Kriz, Alan L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
; TITLE OF INVENTION: IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/080,625  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Hanson, Robert E.  
;; REGISTRATION NUMBER: P-42,628  
;; REFERENCE/DOCKET NUMBER: DEKM:161  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (512) 418-3000  
;; TELEFAX: (512) 474-7577  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3336 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
US-09-080-625-2

Query Match 92.4%; Score 19.4; DB 4; Length 3336;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttgactataacttgac 21  
|||||  
Db 2942 GCTTGGACTATAACTGTGAC 2962

RESULT 12  
US-09-080-625-5  
; Sequence 5, Application US/09080625  
; Patent No. 6307123  
; GENERAL INFORMATION:  
; APPLICANT: Kriz, Alan L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
; TITLE OF INVENTION: IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/080,625  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, Robert E.  
; REGISTRATION NUMBER: P-42,628  
; REFERENCE/DOCKET NUMBER: DEKM:161  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3694 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-080-625-5

Query Match 92.4%; Score 19.4; DB 4; Length 3694;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataaacttgac 21  
|||||

Db 3300 GCTTGGACTATAACTCTGAC 3320

## RESULT 13

US-09-080-625-4  
; Sequence 4, Application US/09080625  
; Patent No. 6307123  
; GENERAL INFORMATION:  
; APPLICANT: Kriz, Alan L.  
; APPLICANT: Spencer, T. Michael  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
; TITLE OF INVENTION: IDENTIFICATION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hanson, Robert E.  
; REGISTRATION NUMBER: P-42,628  
; REFERENCE/DOCKET NUMBER: DEKM.161  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3877 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-080-625-4

Query Match 92.4%; Score 19.4; DB 4; Length 3877;  
Best Local Similarity 95.2%; Pred. No. 0.33;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataaacttgac 21  
|||||

Db 3483 GCTTGGACTATAACTCTGAC 3503

## RESULT 14

US-08-817-188-1/c  
; Sequence 1, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
; EARLIER FILING DATE: 1996-07-31

; EARLIER APPLICATION NUMBER: EP 95401844.6  
; EARLIER FILING DATE: 1995-08-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4946  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pTHW107  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: T-DNA right border (RB)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((97)..(330))  
; OTHER INFORMATION: 3'g7; 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
; OTHER INFORMATION: T-DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
; OTHER INFORMATION: transferase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2658)..(3031))  
; OTHER INFORMATION: 3' nos; 3' untranslated region containing the  
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
; OTHER INFORMATION: gene of Agrobacterium T-DNA  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3032)..(3367))  
; OTHER INFORMATION: barnase: region coding for barnase  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3368)..(4876))  
; OTHER INFORMATION: PTA29; promoter region of TA29 gene of Nicotiana  
; OTHER INFORMATION: tabacum  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((4922)..(4946))  
; OTHER INFORMATION: LB: T-DNA left border  
US-08-817-188-1

Query Match 92.4%; Score 19.4; DB 3; Length 4946;  
Best Local Similarity 95.2%; Pred. No. 0.34;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcttgactataaacttgac 21  
|||||

Db 163 GCTTGGACTATAACTCTGAC 143

## RESULT 15

US-08-817-188-5/c  
; Sequence 5, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366

```
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-LS1 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
```

US-08-817-188-5

```
Query Match          92.4%; Score 19.4; DB 3; Length 5560;
Best Local Similarity 95.2%; Pred. No. 0.34;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcttggaactataactttgac 21
   |||||
Db 150 GCITGGACTATATACCTGAC 130
```

Search completed: February 25, 2002, 18:05:52  
Job time: 18600 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:17 ; Search time 8261.74 Seconds  
(without alignments)  
27.314 Million cell updates/sec

Title: US-09-698-903B-12

Perfect score: 21

Sequence: 1 gcttgactataataacttgac 21

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_estl:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.4	82.9	344	13 AQ117151	AQ117151 HS_2188_A
2	17.4	82.9	633	13 AZ464390	AZ464390 LM0273D15
3	16.8	80.0	291	10 AV035892	AV035892 AV035892
4	16.8	80.0	415	13 AQ172470	AQ172470 HS_3190_B
C 5	16.8	80.0	449	10 AI675738	AI675738 wc40h04.x
6	16.8	80.0	449	11 BE802728	BE802728 sr44g04.y
C 7	16.8	80.0	470	13 AZ133172	AZ133172 OSJNB010
8	16.8	80.0	483	10 AA310103	AA310103 EST180941
9	16.8	80.0	487	13 B74437	B74437 CIT-HSP-202
C 10	16.8	80.0	505	13 AO666336	AO666336 HS_5374_A
11	16.8	80.0	537	13 AQ836967	AQ836967 HS_4553_A
12	16.8	80.0	600	11 BE846195	BE846195 232148 BA

C 13	16.8	80.0	854	13 BH132866	BH132866
C 14	16.8	80.0	968	13 AZ670679	AZ670679
C 15	16.4	78.1	301	10 BB103479	BB103479
16	16.4	78.1	302	10 AA407887	AA407887
C 17	16.4	78.1	305	10 BB117588	BB117588
18	16.4	78.1	309	10 BB235252	BB235252
19	16.4	78.1	310	10 AU021901	AU021901
20	16.4	78.1	353	10 AU021997	AU021997
C 21	16.4	78.1	376	10 AA823300	AA823300
C 22	16.4	78.1	405	10 AA606880	AA606880
23	16.4	78.1	408	10 AU021898	AU021898
24	16.4	78.1	432	10 AA763279	AA763279
25	16.4	78.1	453	13 AQ236085	AQ236085
26	16.4	78.1	470	10 AU017121	AU017121
C 27	16.4	78.1	493	10 AA690887	AA690887
28	16.4	78.1	503	10 AU042554	AU042554
29	16.4	78.1	514	10 AI315372	AI315372
C 30	16.4	78.1	576	10 BE325283	BE325283
31	16.4	78.1	605	10 AW109828	AW109828
C 32	16.4	78.1	613	13 AZ411070	AZ411070
C 33	16.4	78.1	667	13 AQ198178	AQ198178
C 34	16.4	78.1	678	13 AQ201499	AQ201499
35	16.4	78.1	863	11 BF978014	BF978014
36	16.2	77.1	280	10 BB337839	BB337839
C 37	16.2	77.1	305	10 AA013302	AA013302
C 38	16.2	77.1	411	10 AA554565	AA554565
C 39	16.2	77.1	412	11 BF841371	BF841371
C 40	16.2	77.1	423	13 AZ273622	AZ273622
41	16.2	77.1	441	11 H04932	H04932
42	16.2	77.1	445	13 AQ710257	AQ710257
43	16.2	77.1	458	13 AQ493489	AQ493489
44	16.2	77.1	468	10 BE246449	BE246449
45	16.2	77.1	485	13 AG023449	AG023449

#### ALIGNMENTS

RESULT 1	AQ117151/c	AQ117151	344 bp	DNA	GSS	22-SEP-1998
LOCUS	HS_2188_A1_D04_MF	CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2188 Col=7 Row=G, DNA sequence.				
DEFINITION	AQ117151					
ACCESSION	AQ117151					
VERSION	AQ117151.1	GI:3494942				
KEYWORDS	GSS.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.					
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome					
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)					
MEDLINE	99380589					
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence Tagged Connector Plate: 2188 row: G column: 7 Class: BAC ends High quality sequence stop: 344. Location/Qualifiers 1. 344 /organism="Homo sapiens" /db_xref="taxon:9606"					
FEATURES	source					

```

/clone="Plate-2188 Col-7 Row-G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 105 a 53 c 85 g 101 t
ORIGIN

```

```

Query Match 82.9%; Score 17.4; DB 13; Length 344;
Best Local Similarity 94.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cttggactataataacttga 20
| | | | | | | | | | | | | | | | | |
Db 167 CATGGACTATAACTTGA 149

```

```

RESULT 2
AZ464390
LOCUS
DEFINITION
Clone UUGC1M0273D15 R, DNA sequence.
ACCESSION
AZ464390.1 GI:10622515
VERSION
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 633)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0273 row: D column: 15
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 633.
Location/Qualifiers
1. .633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0273D15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number

```

```

Inducible derivative of plasmid pL1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 183 a 133 c 136 g 181 t
ORIGIN

```

```

Query Match 82.9%; Score 17.4; DB 13; Length 633;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cttggactataataacttga 20
| | | | | | | | | | | | | | | | | |
Db 473 CTTGGACTATAAATACTTGA 491

```

```

RESULT 3
AV035892
LOCUS
DEFINITION
AV035892 Mus musculus adult C57BL/6J placenta Mus musculus CDNA
clone 1600016H12, mRNA sequence.
ACCESSION
AV035892
VERSION
AV035892.1 GI:4855557
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 291)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sojabe,Y., Sugahara
,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomioka,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermolabile and its application for the synthesis of full length cDNA
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. .291
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1600016H12"
/clone_lib="Mus musculus adult C57BL/6J placenta"
/sex="female"
/tissue_type="placenta"
/dev_stage="adult"
BASE COUNT 68 a 56 c 66 g 101 t
ORIGIN

```

```

FEATURES
source
1. .633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0273D15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|4732114|gb|AF129072.1), a copy-number

```

```

Query Match 80.0%; Score 16.8; DB 10; Length 291;
Best Local Similarity 90.0%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```



```

Oy 2 cttggactataacttgac 21
    ||||| ||||| ||||| |||||
Db 13 CTTGGACTATATATCGGAC 32

RESULT 4
AQL172470
LOCUS HS_3190_B2_H07_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3190 Col-14 Row-P, DNA sequence. 17-OCT-1998
ACCESSION AQL172470
VERSION AQL172470.1 GI:3569837
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3190 row: P column: 14
Class: BAC ends
High quality sequence stop: 415.
FEATURES
    source
    location/Qualifiers
    1..415
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="CIT Approved Human Genomic Sperm Library D"
    /sex="male"
    /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 77 a 117 c 69 g 151 t 1 others
ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 415;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 cttggactataacttgac 21
    ||||| ||||| ||||| |||||
Db 27 CTTGCAATATATGCTTGAC 46

RESULT 5
A1675738/c
LOCUS A1675738 449 bp mRNA EST 17-DEC-1999
DEFINITION wc40h04.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321143 3' similar to gb:W73255_rnal VASCULAR CELL ADHESION PROTEIN 1
PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION A1675738
VERSION A1675738.1 GI:4876218
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 449)

AUTHORS
TITLE
JOURNAL
COMMENT

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
cDNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert length: 567 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 411.
FEATURES
    source
    location/Qualifiers
    1..449
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:2321143"
    /clone_lib="NCI-CGAP_Pr28"
    /sex="male"
    /dev_stage="adult"
    /lab_host="DH10B"
    /note="Organ: prostate; Vector: pMT3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 178 a 77 c 63 g 131 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 449;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gcttggactataacttga 20
    ||||| ||||| ||||| |||||
Db 419 GCTTGGACTATATATTTA 400

RESULT 6
BE802728/c
LOCUS BE802728 449 bp mRNA EST 20-SEP-2000
DEFINITION sr44904.y1 Gm-c1051 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1051-943 5', mRNA sequence.
ACCESSION BE802728
VERSION BE802728.1 GI:10233840
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 449)
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project

```

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 High quality sequence stop: 402.

## FEATURES

## source

1. .449  
 Location/Qualifiers  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl051-943"  
 /clone\_1lb="Gm-cl051"  
 /tissue\_type="floral meristematic mRNA"  
 /lab\_host="DH108"  
 /note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:  
 XhoI; The cDNA library was constructed from floral  
 meristematic mRNA provided by Dr. Halina Knap of Clemson  
 University. Complementary DNA was synthesized from mRNA  
 using a primer consisting of a poly(dT) sequence with a  
 blunt-ended cDNA fragments followed by XhoI digestion. The  
 cDNA fragments were directionally cloned into the  
 EcoRI-XhoI restriction site of the pBluescript vector. The  
 ligated cDNA fragments were transformed into DH108 host  
 cells (GibcoBRL). This library was constructed in the  
 laboratory of Dr. Randy Shoemaker."  
 BASE COUNT 149 a 73 c 80 g 147 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 449;  
 Best Local Similarity 90.0%; Pred. No. 4.8e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cttggactataactttgac 21  
 ||||| ||||| ||||| |||||  
 Db 358 CTTGGACTATATAAAGTTGAC 339

## RESULT 7

AZ133172  
 LOCUS AZ133172 470 bp DNA GSS 02-JUN-2000  
 DEFINITION OSJNB0108P05r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic  
 clone OSJNB0108P05r, DNA sequence.

ACCESSION AZ133172  
 VERSION AZ133172.1 GI:8212071  
 KEYWORDS GSS.

## SOURCE

## ORGANISM

Oryza sativa.  
 Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 470)  
 Wing, R.A. and Dean, R.A.  
 A BAC End Sequencing Framework to Sequence the Rice Genome  
 Unpublished (1998)  
 Contact: Wing RA

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAAACAGCTAGCACCATG

Class: BAC ends

High quality sequence start: 101

High quality sequence stop: 368.

Location/Qualifiers

## FEATURES

## source

1. .470  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="OSJNB0108P05r"  
 /clone\_1lb="CUGI Rice BAC Library (EcoRI)"  
 /tissue\_type="leaf"  
 /lab\_host="E. coli DH108"

/note="Vector: pBACIndigo; Site\_1: EcoRI; Site\_2: EcoRI;  
 Rice is the most important food crop in the world. Half of  
 the world population, especially those inhabiting highly  
 populated areas of the humid tropics and subtropics, rely  
 on rice as their primary source of carbohydrate.  
 Monocytiledonous rice is a diploid plant (2n=24) with a  
 haploid genome equivalent of 431 Mbp (Arumuganathan and  
 Earle, 1991). The relatively small genome of rice, three  
 times larger than that of Arabidopsis, makes it suitable  
 for genomic studies. In order to facilitate positional  
 cloning, physical mapping and genome sequencing of rice,  
 we have constructed a BAC library from Oryza sativa,  
 Nipponbare variety using EcoRI as the cloning enzyme. The  
 library contains 55,296 clones with an average insert size  
 of 121 Kb providing approximately 15 haploid genome  
 equivalents. The deep coverage allows the isolation a  
 particular sequence with a probability of 99.9 %. Three  
 high density filters, each containing 18,432 clones  
 (doubly spotted), represent the whole library for colony  
 screening and can be requested from the Clemson University  
 BAC/EST Resource Center (www.genome.clemson.edu)."

/lab\_host="E. coli DH108"

/note="Vector: pBACIndigo; Site\_1: EcoRI; Site\_2: EcoRI;

Rice is the most important food crop in the world. Half of

the world population, especially those inhabiting highly

populated areas of the humid tropics and subtropics, rely

on rice as their primary source of carbohydrate.

Monocytiledonous rice is a diploid plant (2n=24) with a

haploid genome equivalent of 431 Mbp (Arumuganathan and

Earle, 1991). The relatively small genome of rice, three

times larger than that of Arabidopsis, makes it suitable

for genomic studies. In order to facilitate positional

cloning, physical mapping and genome sequencing of rice,

we have constructed a BAC library from Oryza sativa,

Nipponbare variety using EcoRI as the cloning enzyme. The

library contains 55,296 clones with an average insert size

of 121 Kb providing approximately 15 haploid genome

equivalents. The deep coverage allows the isolation a

particular sequence with a probability of 99.9 %. Three

high density filters, each containing 18,432 clones

(doubly spotted), represent the whole library for colony

screening and can be requested from the Clemson University

BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 163 a 75 c 94 g 131 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 470;

Best Local Similarity 90.0%; Pred. No. 4.9e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcttgactataacttga 20  
 ||||| ||||| ||||| |||||  
 Db 277 GCTTGGGCTCTAATACTTGA 296

## RESULT 8

## AA310103

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## 1 (bases 1 to 483)

## ADAMS, M.D.,

## C.J.,

## Lee, N.H.,

## Kirkness, E.F.,

## Weinstock, K.G.,

## Fuldner, R.A.,

## Bult

## O.,

## Sutton, G.,

## Blake, J.A.,

## Brandon, R.C.,

## Man-Wai, C.,

## Clayton, R.A.,

## Cline, T.R.,

## Cotton, M.D.,

## Earle-Hughes, J.,

## Fine, L.D.,

## Fitzgerald

## L.M.,

## Fitzhugh, W.M.,

## Fritchman, J.L.,

## Geoghagen, N.S.,

## Glodek, A.,

## Gnehm, C.L.,

## Hanna, M.C.,

## Hedblom, E.,

## Hinkle, P.S.,

## Kelley, J.M.,

## Kelley, J.C.,

## Liu, L.-I.,

## Marmaros, S.M.,

## Merrick, J.M.,

## Moreno-Palanges, R.F.,

## McDonald, L.A.,

## Nguyen, D.T.,

## Pelligrino, S.M.,

## Phillips, C.A.,

## Ryder, S.E.,

## Scott, J.L.,

## Saudek, D.M.,

## Shirley, R.,

## Small, K.V.,

## Spriggs, T.A.,

## Utterback, T.R.,

## Weidman, J.F.,

## Li, Y.,

## Dimkarik, D.P.,

## Cao, L.,

## Cepeda, M.A.,

## Coleman, T.A.,

## Collins, E.J.,

## Dimkarik, D.,

## Feng, D.-F.,

## Ferrie, A.,

## Fischer, C.,

## Hastings, G.A.,

## He, W.W.,

## Hu, J.S.,

## Greene, J.M.,

## Gruber, J.,

## Hudson, P.,

## Kim, A.K.,

## Kozak, D.L.,

## Kunsch, C.,

## Hungjun, J.,

## Li, H.,

## Meissner, P.S.,

## Olsen, H.,

## Raymond, L.,

## Wei, Y.F.,

## Wing, J.,

## Xu, C.,

## Yu, G.L.,

## Ruben, S.M.,

## Dillion, P.J.,

## Fannon

## M.R.,

## Rosen, C.A.,

## Haseltine, W.A.,

## Fields, C.,

## Fraser, C.M.,

## Venter, J.C.

**TITLE** Initial assessment of human gene diversity and expression patterns  
**JOURNAL** based upon 83 million nucleotides of cDNA sequence  
**MEDLINE** Nature 377 (6547 Suppl), 3-174 (1995)  
**COMMENT** 96026280  
 Other ESTs: THC122434  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org

For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

# **FEATURES**

## source

Location/Qualifiers  
 1..483  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):156165"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Jurkat T-cells V"  
 /cell\_type="T-lymphocyte"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 xhoi"

**BASE COUNT** 113 a 142 c 136 g 85 t 7 others

## ORIGIN

**Query Match** 80.0%; Score 16.8; DB 10; Length 483;  
**Best Local Similarity** 85.7%; Pred. No. 4.9e+02;  
**Matches** 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

**Qy** 1 gcttgactataacttgac 21

||||||| ||| |||||  
**Db** 123 GCTTGGACGAATTCITGAC 143

## RESULT

9

**B74437** 487 bp DNA GSS 24-OCT-1998  
**LOCUS** CIT-HSP-2028F18.TF CIT-HSP Homo sapiens genomic clone 2028F18, DNA  
**DEFINITION** sequence.

**ACCESSION** B74437

**VERSION** B74437.1 GI:2770124

**KEYWORDS** GSS.

**SOURCE** human.

**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** 1 (bases 1 to 487)  
**AUTHORS** Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden  
 ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.  
 and Venter,J.C.

**TITLE** Use of a random BAC End Sequence Database for Sequence-Ready Map

Building

**JOURNAL** Unpublished (1997)

**COMMENT** Other\_GSSs: CIT-HSP-2028F18.TR

Contact: Mark Adams

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics ([inforesgen.com](http://inforesgen.com)). BAC  
 end search page:

[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)

Seq primer: M13-21

Class: BAC ends.

**FEATURES** Location/Qualifiers

## source

1..487  
 /organism="Homo sapiens"

/db\_xref="GDB:7048076"  
 /db\_xref="taxon:9606"  
 /clone="2028F18"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
 HindIII"

**BASE COUNT** 160 a 101 c 100 g 126 t

## ORIGIN

**Query Match** 80.0%; Score 16.8; DB 13; Length 487;  
**Best Local Similarity** 90.0%; Pred. No. 4.9e+02;  
**Matches** 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**Qy** 2 cttggactataacttgac 21

||||||| ||||| ||  
**Db** 57 CTTGGACTGTAATCTTGTG 76

## RESULT

10

**AQ666336/c**

**LOCUS**

**DEFINITION**

**ACCESSION**

**VERSION**

**KEYWORDS**

**SOURCE**

**ORGANISM**

**REFERENCE**

**AUTHORS**

**TITLE**

**JOURNAL**

**MEDLINE**

**COMMENT**

**SEQUENCE-TAGGED CONNECTORS:** A sequence approach to mapping and  
 scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from

BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))

or from Resear h Genetics ([inforesgen.com](http://inforesgen.com)). BAC end Web Server:

<http://www.htsc.washington.edu>

Plate: 950 row: O column: 1

Seq primer: T7

Class: BAC ends

High quality sequence stop: 505.

**FEATURES** Location/Qualifiers

source

1..505

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-950 Col-1 Row-O"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the

pBACE3.6 vector at EcoRI sites"

**BASE COUNT** 141 a 121 c 85 g 154 t

**ORIGIN**

```

Query Match      80.0%; Score 16.8; DB 13; Length 505;
Best Local Similarity 90.0%; Pred. No. 4.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactctgac 21
||||| ||| ||||| |||
Db 500 CTTGGAGTATTATACCTTCAC 481

RESULT 11
A0836967      537 bp      DNA      GSS      30-AUG-1999
DEFINITION   HS_4553_A1_D09_T7A CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=4553 Col=17 Row=G, DNA sequence.
ACCESSION   A0836967
VERSION     A0836967.1 GI:5806929
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 537)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 4553 row: G column: 17
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 537.
FEATURES     source
            Location/Qualifiers
                1..537
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="Plate=4553 Col=17 Row=G"
                /sex="male"
                /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH108"
BASE COUNT   154 a 128 c 115 g 137 t 3 others
ORIGIN

Query Match      80.0%; Score 16.8; DB 13; Length 537;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactctgac 21
||||| ||| ||||| |||
Db 438 CTTGGACTAGTACTCTTAC 457

RESULT 12
BE846195      600 bp      mRNA      EST      25-SEP-2000
DEFINITION   BE846195 232148 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE846195
VERSION     BE846195.1 GI:10283019
KEYWORDS    EST.
SOURCE      cow.

Query Match      80.0%; Score 16.8; DB 11; Length 600;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactctgac 21
||||| ||| ||||| |||
Db 532 CTTGGACTCTATCTCTTGAC 551

RESULT 13
BH132866/c    854 bp      DNA      GSS      07-AUG-2001
LOCUS        ENTNI55TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
DEFINITION   BH132866.1 GI:15091927
ACCESSION   BH132866
VERSION     BH132866.1
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Entamoeba histolytica.
REFERENCE   1 (bases 1 to 854)
AUTHORS     Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE       Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: enta@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

```

```

ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE     1 (bases 1 to 600)
AUTHORS      Sonstegard,T.S., Capuco,A.V., Van Tassel,C.F., Ashwell,M.S. and Wells,K.D.
TITLE        Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library
JOURNAL      Unpublished (2000)
COMMENT      Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@psi.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
FEATURES     source
            Location/Qualifiers
                1..600
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone_lib="BARC 5BOV"
                /tissue_type="pooled"
                /lab_host="DH108"
                /note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."
BASE COUNT   150 a 140 c 112 g 198 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 11; Length 600;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cttggactataactctgac 21
||||| ||| ||||| |||
Db 532 CTTGGACTCTATCTCTTGAC 551

RESULT 13
BH132866/c    854 bp      DNA      GSS      07-AUG-2001
LOCUS        ENTNI55TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
DEFINITION   BH132866.1 GI:15091927
ACCESSION   BH132866
VERSION     BH132866.1
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Entamoeba histolytica.
REFERENCE   1 (bases 1 to 854)
AUTHORS     Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE       Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: enta@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

```

Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 18  
 High quality sequence stop: 742.  
 Location/Qualifiers

## FEATURES

source

1. .854  
 /organism="Entamoeba histolytica"  
 /strain="HM1:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOSt; Site\_1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Borell, Oxford University Press, 1999)."

BASE COUNT 251 a 100 c 95 g 408 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 854;  
 Best Local Similarity 90.0%; Pred. No. 5.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcttggaactataacttga 20  
 ||||| ||||| ||||| |||||  
 Db 143 GCTTGGACTAAATATTGA 124

RESULT 14  
 LOCUS ENT3479 968 bp DNA 14-DEC-2000  
 DEFINITION Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 genomic, DNA sequence.  
 ACCESSION Z670679  
 VERSION Z670679.1 GI:11807825  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica.  
 Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 968)  
 Loftus, B., Van Aken, S. and Fraser, C.  
 Determination of clone end sequences from Entamoeba histolytica  
 HM1:IMSS sheared DNA library  
 Unpublished (2000)  
 JOURNAL Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: enta@tigr.org  
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared  
 DNA library

Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 94  
 High quality sequence stop: 406.  
 Location/Qualifiers

## FEATURES

source

1. .968  
 /organism="Entamoeba histolytica"  
 /strain="HM1:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOSt; Site\_1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Borell, Oxford University Press, 1999)."

BASE COUNT 254 a 127 c 210 g 377 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 968;  
 Best Local Similarity 90.0%; Pred. No. 5.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gcttggaactataacttga 20  
 ||||| ||||| ||||| |||||  
 Db 363 GCTTGGACTAAATATTGA 344

RESULT 15

LOCUS

BB103479 301 bp mRNA EST 27-JUN-2000  
 BB103479 RIKEN full-length enriched, 12 days embryo, embryonic body  
 between diaphragm region and neck Mus musculus cDNA clone  
 9430087K10 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BB103479 301 bp mRNA EST 27-JUN-2000  
 BB103479 RIKEN full-length enriched, 12 days embryo, embryonic body  
 between diaphragm region and neck Mus musculus cDNA clone  
 9430087K10 3', mRNA sequence.  
 BB103479.1 GI:8756047  
 EST.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 301)  
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
 Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata  
 Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya  
 T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,  
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino  
 M., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Konno, H., et al.)  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

TITLE

JOURNAL

COMMENT

Email: genome-res@gsc.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki  
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Thermostabilization and thermoactivation of thermostable enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
 Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,





```
BASE COUNT      7 a      6 c      6 g      3 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 1 AACGAGTGTCTAGCTAGACCAGC 22

RESULT 2
AX172481
LOCUS      AX172481      22 bp      DNA      PAT      03-JUL-2001
DEFINITION Sequence 42 from Patent WO0141558.
ACCESSION  AX172481
VERSION     AX172481.1 GI:14597593
KEYWORDS   .
SOURCE      synthetic construct.
ORGANISM    artificial sequence.
REFERENCE   1 (bases 1 to 22)
AUTHORS    de Both,G. and de Beuckeleer,M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL    Patent: WO 0141558-A 42 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
            1..22
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="primer BNA05"
            /note="6 c      6 g      3 t"

BASE COUNT      7 a      6 c      6 g      3 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 1 AACGAGTGTCTAGCTAGACCAGC 22

RESULT 3
BNACRUC
LOCUS      BNACRUC      1624 bp      mRNA      PLN      27-APR-1993
DEFINITION Rapeseed cruciferin mRNA, complete cds.
ACCESSION  M16860
VERSION     M16860.1 GI:167133
KEYWORDS   12S storage protein; cruciferin.
SOURCE      Rapeseed (cultivar Tower) embryo (26-day post anthesis), cDNA to
            mRNA, clone pC1.
ORGANISM    Brassica napus.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
            1 (bases 1 to 1624)
            Simon,A.E., Tenbarger,K.M., Scofield,S.R., Finkelstein,R.R. and
            Crouch,M.L.
            Nucleotide sequence of a cDNA clone of Brassica napus 12S storage
            protein shows homology with legumin from Pisum sativum
            Plant Mol. Biol. 5, 191-201 (1985)
            Location/Qualifiers
            1..1624
            /organism="Brassica napus"
            /db_xref="taxon:3708"
            <1..1624
            /product="CRUC mRNA"
            17..85
            /note="cruciferin signal peptide"

CDS
17..1483
/note="cruciferin precursor"
/codon_start=1
/protein_id="AAA32988.1"
/db_xref="GI:167134"
/translation="MARLSSLLSFLALIFLHGSTAQQFPNECQLDQLNALEPSSHVL
KAEAGRIEVDHHPOLRCGSGVSVRYIIIESKGLYLPSPFSTAKLSFVAKGGLMGRV
VLCAETFDSSVFPSPGSGSPFGEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG
FRDMHOKVEHIRTGTDTIHPGVAQWPFYNDGNQPLVIYVSLASHONQIDNRNPFY
LAGNPPQGVWIEGREQQPKNILNGFTPEVLAKAFKIDVTAQQQLNQDNRGNLIIR
VOGPFVIRPPLRSQRPOEEVNGLEETICSARCTDNLDPPSNADYKPKQGYISTLNS
YDLPIRLFLSALRGSIQANAMVLPQMANANAVLYVTDGEAHVQVAVNDGDRVFDG
QVSGQLLSIQGFSVVKRATSEQFWIEFKTNAQININTLAGRTSVLRGLPLEVISN
GYOISLEEARVKENTTETLTTHSSGPASYGGPRKADA"
mat_peptide      86..910
/note="cruciferin alpha subunit"
mat_peptide      911..1480
/note="cruciferin beta subunit"
BASE COUNT      457 a      425 c      387 g      355 t
ORIGIN      70 bp upstream of NcoI site.

Query Match      100.0%; Score 22; DB 8; Length 1624;
Best Local Similarity 100.0%; Pred. No. 0.098;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 98 AACGAGTGTCTAGCTAGACCAGC 119

RESULT 4
I24540
LOCUS      I24540      3113 bp      DNA      PAT      07-OCT-1996
DEFINITION Sequence 20 from patent US 5543576.
ACCESSION  I24540
VERSION     I24540.1 GI:1604410
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 3113)
AUTHORS    van Ooijen,A.J., Rietveld,K., Hoekema,A., Pen,J., Sijmons,P.C.,
            Verwoerd,T.C. and Quax,W.J.
TITLE      Production of enzymes in seeds and their use
JOURNAL    Patent: US 5543576-A 20 06-AUG-1996;
FEATURES   Location/Qualifiers
            source
            1..3113
            /organism="unknown"
            /db_xref="taxon:32630"
            /note="6 c      6 g      881 t"

BASE COUNT      961 a      685 c      586 g      881 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 3113;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
|||||
Db 790 AACGAGTGTCTAGCTAGACCAGC 811

RESULT 5
I83673
LOCUS      I83673      3113 bp      DNA      PAT      10-AUG-1998
DEFINITION Sequence 2 from patent US 5714474.
ACCESSION  I83673
VERSION     I83673.1 GI:3407203
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 3113)
```



AUTHORS Van Ooljen,A.J.J., Rietveld,K., Hoekema,A., Pen,J.,  
Sijmons,P.Christian, Verwoerd,T.Cornelis and Quax,W.Johannes.  
TITLE Production of enzymes in seeds and their use  
JOURNAL Patent: US 5714474-A 2 03-FEB-1998;  
FEATURES Location/Qualifiers  
source 1..3113  
/organism="unknown"  
BASE COUNT 961 a 684 c 586 g 882 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
|||||  
Db 790 AACGAGTGTCTAGACCAGC 811

RESULT 6  
BNCRUA BNCRUA 3113 bp DNA PLN 10-FEB-1999  
LOCUS Brassica napus cruA gene for cruciferin.  
DEFINITION  
ACCESSION X14555  
VERSION X14555.1 GI:117810  
KEYWORDS cruA gene; cruciferin; seed storage protein.  
SOURCE  
ORGANISM  
Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS Ryan,A.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAR-1989) Ryan A.J., Department of Biological  
Sciences, University of Durham, South Road, Durham, DH1 3LE, United  
Kingdom  
REFERENCE 2 (bases 1 to 3113)  
AUTHORS Ryan,A.J., Royal,C.L., Hutchinson,J. and Shaw,C.H.  
TITLE Genomic sequence of a 12S seed storage protein from oilseed rape  
(Brassica napus c.v. jet neuf)  
JOURNAL Nucleic Acids Res. 17 (9), 3584 (1989)  
MEDLINE 89263796  
FEATURES  
source  
1. .3113  
Location/Qualifiers  
/organism="Brassica napus"  
/strain="Jet Neuf"  
/db\_xref="taxon:3708"  
/tissue\_type="leaf"  
644. .650  
promoter  
/note="pot. TATA-box"  
precursor\_RNA 580. .3010  
/note="primary transcript"  
CDS join(709. .991,1220. .1581,2049. .2870)  
/codon\_start=1  
/product="cruciferin"  
/protein\_id="CAA32692.1"  
/db\_xref="GI:17811"  
/db\_xref="SWISS-PROT:P11090"  
/translation="MARLSLLSFLALLTLFHGSTAQOQFPNECQLDQLNALEPSHVL  
KAEAGRIEVDHAPOLRCSGVSFVRYIESKGLYLPSEFFSTAKLSFVAKGEGLMGRV  
VPGCAETFDQSVFPGSGSPGEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG  
FRDMHQKVEHRTGDTATHPGVAQWYNDGNQPLVIVSLDLASHQNLDRNRPFY  
LQGNPQGVWIEGREQQPNKILNGEFTPEVLAKAFKIDVTKAQQLNQDNRGNILR  
VQGFPSVIRPLRSRQPEEVNGLEETICSARCTNDLDDPSNADYKPOLGISTLS  
YDLPLRLSLRSLSIRQNAVLPQWNNANAVLYVTDGEAHVQVWVNDGRVFDG  
QVSGQLLSIPOGFSVVKRATSEQFRWIEFKTNANAOINTLAGRTSVLRGLPLEVIS  
GYQISLEARRVKFTIETTLTHSSGPASVGGPRKADA"  
992. .1219  
intron  
/note="intron I"  
intron 1582. .2048  
/note="intron II"

misc\_feature 2915. .2920  
/note="polyA signal"  
misc\_feature 2989. .2994  
/note="polyA signal"  
BASE COUNT 961 a 685 c 586 g 881 t  
ORIGIN

Query Match 100.0%; Score 22; DB 8; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
|||||  
Db 790 AACGAGTGTCTAGACCAGC 811

RESULT 7  
BNC1G BNC1G 3198 bp DNA PLN 04-APR-1995  
LOCUS B.napus Bnc1 gene for cruciferin storage protein.  
DEFINITION  
ACCESSION X59294  
VERSION X59294.1 GI:17790  
KEYWORDS cruciferin; cruciferin storage protein.  
SOURCE  
ORGANISM  
Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 3198)  
AUTHORS Breen,J.P. and Crouch,M.L.  
TITLE Molecular analysis of a cruciferin storage protein gene family of  
Brassica napus  
JOURNAL Plant Mol. Biol. 19 (6), 1049-1055 (1992)  
MEDLINE 92379259  
REFERENCE 2 (bases 1 to 3198)  
AUTHORS Breen,J.P.  
TITLE Direct Submission  
JOURNAL Submitted (24-APR-1991) Breen J.P., Jordan Hall, Indiana  
University, Bloomington, IN 47401, USA  
COMMENT See also M16860 & X59295 (for Bnc2 gene).  
FEATURES  
Location/Qualifiers  
1. .3198  
/organism="Brassica napus"  
/strain="C.v. Tower"  
/db\_xref="taxon:3708"  
/cell\_line="ED8767"  
/clone\_lib="Ch4a"  
/clone="lambda Bnc1"  
join(<709. .991,1220. .1584,2069. .2488,2653. .>3057)  
/gene="Bnc1"  
709. .3057  
/gene="Bnc1"  
join(709. .991,1220. .1584,2069. .2488,2653. .3057)  
/gene="Bnc1"  
/codon\_start=1  
/product="cruciferin storage protein"  
/protein\_id="CAA41984.1"  
/db\_xref="GI:762919"  
/db\_xref="SWISS-PROT:P33523"  
/translation="MARLSLLSFLALLTLFHGSTAQOQFPNECQLDQLNALEPSHVL  
KAEAGRIEVDHAPOLRCSGVSFVRYIESKGLYLPSEFFSTAKLSFVAKGEGLMGRV  
VPGCAETFDQSVFPGSGSPGEGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOGOG  
FRDMHQKVEHRTGDTATHPGVAQWYNDGNQPLVIVSLDLASHQNLDRNRPFY  
LQGNPQGVWIEGREQQPNKILNGEFTPEVLAKAFKIDVTKAQQLNQDNRGNILR  
VQGFPSVIRPLRSRQPEEVNGLEETICSARCTNDLDDPSNADYKPOLGISTLS  
YDLPILRLSLRSLSIRQNAVLPQWNNANAVLYVTDGEAHVQVWVNDGRVFDG  
QVSGQLLSIPOGFSVVKRATSEQFRWIEFKTNANAOINTLAGRTSVLRGLPLEVI  
SNGQISLEARRVKFTIETTLTHSSGPASVGGPRKADA"  
<709. .991  
/gene="Bnc1"  
/number=1  
992. .1219  
exon  
intron

```

/gene="BnCl"
/number=1
1220..1584
/gene="BnCl"
/number=2
1585..2068
/gene="BnCl"
/number=2
2069..2488
/gene="BnCl"
/number=3
2489..2652
/gene="BnCl"
/number=3
2654..>3057
/gene="BnCl"
/number=4
3102..3107
polyA_signal 3107..3107
polyA_signal 3176..3181
BASE COUNT 992 a 690 c 600 g 916 t
ORIGIN

```

```

Query Match 100.0%; Score 22; DB 8; Length 3198;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
|||||
Db 790 AACGAGTGTCTAGCTAGACCAGC 811

```

```

RESULT 8
AL139402
LOCUS
DEFINITION Human DNA sequence from clone RP4-808F24 on chromosome
Xp11.23-11.4. Contains STSS and GSSs, complete sequence.
ACCESSION AL139402.11 GI:9944136
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 10959)
Clark, S.
Direct Submission
Submitted (24-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 29, 2000 this sequence version replaced gi:9926454.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP4-808F24 is from the library RPCI-4 constructed at the Roswell

```

Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>  
 VECTOR: pCYPAC2  
 IMPORTANT: This sequence is not the entire insert of clone RP4-808F24 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP4-774G10 is at 10860 in this sequence.  
 The true right end of clone RP6-227L5 is at 100 in this sequence.

```

FEATURES
    source
        1..10959
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="X"
            /map="p11.23-11.4"
            /clone="RP4-808F24"
            /clone_lib="RPCI-4"
            /complement(1..87)
            /note="match: STS: Em:HS227L5T"
            2314..2355
            /note="21 copies 2 mer gg 76% conserved"
            2339..2342
            /note="unidirectional dGTP"
            /complement(2797..3221)
            /note="match: GSS: Em:B83505"
            5384..5631
            /note="L1MB8 repeat: matches 5924..6175 of consensus"
            5682..5791
            /note="L1MB8 repeat: matches 5654..5773 of consensus"
            6792..6940
            /note="LTR37A repeat: matches 268..420 of consensus"
            6941..6978
            /note="19 copies 2 mer ta 92% conserved"
            6980..7218
            /note="LTR37A repeat: matches 47..282 of consensus"
            7440..7486
            /note="LTR37A repeat: matches 1..47 of consensus"
            8540..8922
            /note="MSTA repeat: matches 1..426 of consensus"
            /complement(8828..9278)
            /note="match: GSS: Em:AQ223257"
            /complement(8948..9266)
            /note="match: GSS: Em:B53441"
            9300..9420
            /note="MER5B repeat: matches 1..128 of consensus"
            /complement(9616..10040)
            /note="match: GSS: Em:AQ369714"
            10048..10487
            /note="L1MB3 repeat: matches 5455..5901 of consensus"
            10509..10820
            /note="LTR26 repeat: matches 291..600 of consensus"
            10847..10958
            /note="LTR26 repeat: matches 201..320 of consensus"
BASE COUNT 3380 a 2058 c 2220 g 3301 t
ORIGIN

```

```

Query Match 88.2%; Score 19.4; DB 9; Length 10959;
Best Local Similarity 95.2%; Pred. No. 3.7;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 21
|||||
Db 7197 AAGGAGTGTCTAGCTAGACCAG 7217

RESULT 9
ATCCAL
LOCUS
DEFINITION Arabidopsis CRAL gene for L2S seed storage protein.
ACCESSION X14312
VERSION X14312.1 GI:16231
KEYWORDS CRAL gene; seed storage protein; storage protein.
SOURCE thale cress.

```

ORGANISM	Arabidopsis thaliana	ORGANISM	eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae; Arabidopsis.	REFERENCE	1 (bases 1 to 2346)
AUTHORS	Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.	AUTHORS	Pang,P.P., Pruitt,R.E. and Meyerowitz,E.M.
TITLE	Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana	TITLE	Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana
JOURNAL	Plant Mol. Biol. 11, 805-820 (1988)	JOURNAL	Plant Mol. Biol. 11, 805-820 (1988)
FEATURES	Location/Qualifiers	FEATURES	Location/Qualifiers
source	1..2346	source	1..2346
	/organism="Arabidopsis thaliana"		/organism="Arabidopsis thaliana"
	/cultivar="Landsberg erecta"		/db_xref="taxon:3702"
promoter	138..147		/db_xref="GI:166676"
exon	196..496		/translation="MARVSSLSPCLTLLILFHGYAAQGGQGGQFPNECOLDLQNAL EPSHVLKSEAGRIEVDHAPQLRCGVSFARYLIESKGLYLPSFFNTAKLSFVAKGR GLMGKVPICGAETFDQSEFPREGOGSORFDMHQKVEHIRSGDTIATPCVAQM FYNDGQPLIVSVFDLASHQNDLRNPRFYLAGNPPQGVWLOGREQPQKNIFNG FGPEVIAQALKIDLTAAQLQNDNRNIVRVQGFVIRPRLRGORPQEEEEGR HGRHGNGLEETICSACTDNLDDPSRADVYKPOLGYISTLNSYDLPLRFLRSALRG SIRONAMVLPQWNAANAILYETDGEAQIQLVNDGNRVFDGQVSQGLIAPVQGFV VKRATSNRFQWVEKTNANAOINTLAGRTSVLRGLPLEVITNGFQISPEARRVKFNT LETTLTHSSGPASYGRPRVAAA"
	/note="TATA-box"		/note="12S storage protein CRA1"
	/number=1		/number=1
CDS	join(196..496,623..888,985..1431,1547..1951)		497..622
	/product="12S seed storage protein"		/note="12S storage protein CRA1 Intron A"
	/protein_id="CAA32493.1"		623..888
	/db_xref="GI:808936"		/number=2
	/translation="MARVSSLSPCLTLLILFHGYAAQGGQGGQFPNECOLDLQNAL EPSHVLKSEAGRIEVDHAPQLRCGVSFARYLIESKGLYLPSFFNTAKLSFVAKGR GLMGKVPICGAETFDQSEFPREGOGSORFDMHQKVEHIRSGDTIATPCVAQM FYNDGQPLIVSVFDLASHQNDLRNPRFYLAGNPPQGVWLOGREQPQKNIFNG FGPEVIAQALKIDLTAAQLQNDNRNIVRVQGFVIRPRLRGORPQEEEEGR HGRHGNGLEETICSACTDNLDDPSRADVYKPOLGYISTLNSYDLPLRFLRSALRG SIRONAMVLPQWNAANAILYETDGEAQIQLVNDGNRVFDGQVSQGLIAPVQGFV VKRATSNRFQWVEKTNANAOINTLAGRTSVLRGLPLEVITNGFQISPEARRVKFNT LETTLTHSSGPASYGRPRVAAA"		889..984
	/db_xref="SWISS-PROT:P15455"		/number=2
	/number=1		/note="12S storage protein CRA1 Intron B"
intron	497..622		985..1431
exon	623..888		1432..1546
intron	889..984		/number=3
exon	985..1431		1547..1951
intron	1432..1546		/number=4
exon	1547..1951		/note="12S storage protein CRA1 Intron C"
polyA_signal	2015..2020		1547..>1951
polyA_signal	2085..2090		/note="12S storage protein CRA1"
BASE COUNT	714 a 528 c 472 g 632 t	BASE COUNT	714 a 528 c 472 g 632 t
ORIGIN		ORIGIN	
Query Match	85.5%; Score 18.8; DB 8; Length 2346;	Query Match	85.5%; Score 18.8; DB 8; Length 2346;
Best Local Similarity	90.9%; Pred. No. 7.8;	Best Local Similarity	90.9%; Pred. No. 7.8;
Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 aacgagtgtcagctagaccagc 22 	Qy	1 aacgagtgtcagctagaccagc 22 
Db	295 AACGAGTGCACGCTCGACCAGC 316	Db	295 AACGAGTGCACGCTCGACCAGC 316
RESULT	11	RESULT	11
BNC2G		BNC2G	
LOCUS	BNC2G 3081 bp DNA	LOCUS	BNC2G 3081 bp DNA
DEFINITION	B.napus Bnc2 gene for cruciferin storage protein.	DEFINITION	B.napus Bnc2 gene for cruciferin storage protein.
ACCESSION	X59295	ACCESSION	X59295
VERSION	X59295.1 GI:17791	VERSION	X59295.1 GI:17791
KEYWORDS	cruciferin; cruciferin storage protein.	KEYWORDS	cruciferin; cruciferin storage protein.
SOURCE	rape.	SOURCE	rape.
ORGANISM	Brassica napus	ORGANISM	Brassica napus
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae; Brassica.		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae; Brassica.
REFERENCE	1 (bases 1 to 3081)	REFERENCE	1 (bases 1 to 3081)
AUTHORS	Breen,J.P. and Crouch,M.L.	AUTHORS	Breen,J.P. and Crouch,M.L.
TITLE	Molecular analysis of a cruciferin storage protein gene family of Brassica napus	TITLE	Molecular analysis of a cruciferin storage protein gene family of Brassica napus
JOURNAL	Plant Mol. Biol. 19 (6), 1049-1055 (1992)	JOURNAL	Plant Mol. Biol. 19 (6), 1049-1055 (1992)
MEDLINE	92379259	MEDLINE	92379259
REFERENCE	2 (bases 1 to 3081)	REFERENCE	2 (bases 1 to 3081)
AUTHORS	Breen,J.P.	AUTHORS	Breen,J.P.
TITLE	Direct Submission	TITLE	Direct Submission

JOURNAL	Submitted (24-APR-1991) Breen J.P., Jordan Hall, Indiana University, Bloomington, IN 47401, USA
COMMENT	See also M16860 & X59294 (for Bnc1 gene).
FEATURES	Location/Qualifiers
source	1..3081
	/organism="Brassica napus"
	/strain="C.v. Tower"
	/db_xref="taxon:3708"
	/cell_line="ED8767"
	/clone_lib="Ch4a"
	/clone="lambda Bnc2"
mRNA	join(<665..947,1320..1701,1777..2197,2381..>2785)
gene	/gene="Bnc2"
	665..2785
	/gene="Bnc2"
	join(665..947,1320..1701,1777..2197,2381..2785)
CDS	/gene="Bnc2"
	/codon_start=1
	/product="cruciferin storage protein"
	/protein_id="CAA41985.1"
	/db_xref="GI:762920"
	/db_xref="SWISS-PROT:P33524"
	/translation="MARLSLLYFSITVLIFLHGSTAQQFPNECQLDQLNALPESIVL KAPAGIEVDHAPQLRSGSVSFYRIIESGLILPSELTANVSFVAKGGQLMGRV VPGCATFTDSSVFQPGSGPFGEGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ QGGQGGFDMHOKVBIHNSGDTIATHPGVAQFYNNGNPLIVIVAMDLSHQNDL RPSQFLAGNPGQSWLHGRGQQPNINLNGFSPEVLAQAFKIDVTPAQQLNQD NRGNIYVGGPFGVIRPPLKSQRPQETEAANGLEETICSARCTDNLDPNSADVYKPOL GTISILNSVDLPLRLVRLSALRGSIRQNAVLPQWKSNAVLVYTDGEAIOIVND NGDRVDPGVQSQGLLSIPQGSFVVKRATSDQFRWIEFTKTNAAQINTLAGRTSWRG LPLEVTANGYQISLEEARVKENTITLTHSGSPASYGRPRKADA"
exon	<665..947
	/gene="Bnc2"
	/number=1
Intron	948..1319
	/gene="Bnc2"
	/number=1
exon	1320..1701
	/gene="Bnc2"
	/number=2
Intron	1702..1776
	/gene="Bnc2"
	/number=2
exon	1777..2197
	/gene="Bnc2"
	/number=3
Intron	2198..2380
	/gene="Bnc2"
	/number=3
exon	2382..>2785
	/gene="Bnc2"
	/number=4
polyA_signal	2921..2926
BASE COUNT	957 a 623 c 601 g 900 t
ORIGIN	
Query Match	85.5%; Score 18.8; DB 8; Length 3081;
Best Local Similarity	90.9%; Pred. No. 7.9;
Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 aacgagtgctcagctagaccagc 22 
Db	746 AACGAGTGCCAACTAGACCAGC 767
RESULT 12	
AB005239/c	
LOCUS	AB005239 84544 bp DNA PLN 27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MLN1.
ACCESSION	AB005239 BA000015
VERSION	AB005239.1 GI:2264311
KEYWORDS	
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui pl
ORGANISM	clone:MLN1.
	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE	1 (sites)
AUTHORS	Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M., Miyajima,N. and Tabata,S.
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence features of the 1.6 Mb regions covered by twenty physically assigned pl clones
JOURNAL	DNA Res. 4 (3), 215-230 (1997)
MEDLINE	97471969
REFERENCE	2 (bases 1 to 84544)
AUTHORS	Nakamura,Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUL-1997) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=MLN1 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://complib.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremml.zool.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MRH10 and the 3' clone is K9L2.
FEATURES	Location/Qualifiers
source	1..84544
	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/chromosome="5"
	/clone="MLN1"
	/clone_lib="Mitsui pl"
	complement(join(1836..1976,2184..2309,2411..2567, 2999..3108,3221..3368,3666..3807,4158..4248,4368..4473, 4925..5024,5127..5232,5584..5823,6332..6481))
	/note="gb RAD45158.1
	gene_id:MLN1.1
	/codon_start=1
	/evidence=not_experimental
	/product="protein phosphatase 2A 62 kDa B'' regulatory subunit"
	/protein_id="BAB10976.1"
	/db_xref="GI:9759510"
	/translation="MEIDGNDVQLDPELLQLPGLSPVSKENPHIAEELFSQWLSL PETGRVLKSLIDDTKSTPTKSVKPSKSLNACGSLPVSFLNPSRPGSPGSP RFSRQKTSPLQSKVREPQRQLIPQFYFQHGRRPAKRLREQCISAVDQFSNYID GLHMDFKSITKEVKLPFLSSVFLSVIDTSGTIVTRDAF IKYVDGHMLANDVAS QIYNILRQPGCKYLROADFKPVLDELLTHPGLEFLRNTPEQERYAETVIRFYII NRSGTCITLRELKGNLITAMQVDEDDINKVIRFSYEHFVYICRFWELDGDHD FLIKENLKYGNHALTIRVDRIFSQQPRAKTSKVEGMSYEDFAFYLAEDKSS PSLEYWFKCIDLDGCVITPMNQFFVEEQHRECTIQEPLFEDILCQIFDKKE KENCITLQDLKASLGNINFLNLFKMAFETRDPLIRQERENPTLTEDWDRFAQR EYVRLSMEEDVEEVSNGSADVWDEPLSPF"
	complement(join(8026..8331,8417..8550,8651..8714, 8807..8892,8990..9116,9214..9298,9390..9453,9539..9600, 9708..9778,9886..9981,10084..10232,10354..10423,

```
10646..10686,10782..10857))
/notes="gene_id:MLN1.2"
/codon_start=1
/evidence="not_experimental"
/product="casein kinase I"
/protein_id="BAB10977.1"
/db_xref="GI:9759511"
/translation="MDLVIGGKFKLGGKIGSGSGELYLGNNVOTGEBVAVKLENVKT
KHPQLHYESKLYMLQGGSGIPNIKWFGEVDSVMVIDLGPSSLEDLFNCNRKLT
KTVLMLADQLNRFVEMHTGRFLHRDKIPDNFLMGLGRKANQVYIIDFGLGKKYRDQLQ
TKHKIPYRKNKLTGTARYASVNTLGLVEQSRDDLESGLGYVLMFLKGLSPWQGLKA
GTKKQYDRLSSEKSTPIEVLCKNOPSEFVSYPHYCRSLREFDDKPDYSYLKRLFRDI
FIREGYQFYVDMTVLYKYPQIGSSGSSSTRHHTTAKPGFNADPIERQERILCKET
SRKVAASRRPSAGPTTSSRPNDRSRNSDDGPFSSKQTHGDSERANSSSRFRASS
TRYKAAVSSRRPSAGPTTSSRPNDRSRNSDDGPFSSKQTHGDSERANSSSRFRASS
SFSRATAARNTREDQLRSFELLSLRK"
complement(join(12794..13054,13133..13219,13304..13440,
13520..13571,13656..13816,14094..14244))
/notes="gb|AAD20643.1
gene_id:MLN1.3"
/codon_start=1
/evidence="not_experimental"
/product="NBD-like protein"
/protein_id="BAB10978.1"
/db_xref="GI:9759512"
/translation="MAEKNASAVDGAIRVSGMQFSYDVQDPFFDFNLDLPAGSRCLL
VGANGSKTTLTKLTLAGKHMVGKNVQVLDORSAFHDTLVCSDGLSYLGGSWSKTAG
SAGDIPQGGSAFHMTFGVEGIDPFREKILDLDDINLQWRMHKVSDDGQRRRVQICM
GLHPFKFVLDDVTVLDVVARMDLLEFPKECEQRGATYVYATHIPFDGLTWSHL
AYINGGKLLUSAKLDEIKDLTKSPNLVSVEAWLRSETKKEKKTKPKPVVTSPPMSSR
OMAYR"
complement(join(15032..15436,15552..15998,16129..16394,
16518..16818))
/notes="gene_id:MLN1.4"
/codon_start=1
/evidence="not_experimental"
/product="legumin-like protein"
/protein_id="BAB10979.1"
/db_xref="GI:9759513"
/translation="MARVSSLSFCLTLILFHGYAAQGGQGGQFNEQGLDOLNAL
EPHVLKSEAGRIEVDHAPQLRCGVSFARYILLESKGLYLPFSFNTAKLSFVAKGR
GLMKVLPGAETFDQSEFPQREGQVSFARYILLESKGLYLPFSFNTAKLSFVAKGR
FYNDGQPLVYVFDIAHQNDLRNRPFYLAGNQPGQVQLGQREQQPKNFING
FGVEYIAALKIDIQALQONQDNRNIVRVQPGFVRPPLRGQRPQEEEEEER
HGRGNLEETICSACTDNLDDPSRADYKPOLGYISTLNSYDLPILTRLISALRG
SIRQNAVLQWANAANAILYVTDGEAQIOIVDNGNREVEDGOVSGQLTAVPQFSV
VKRATSRFQWVEKTNANAQINTLAGRTSVLRGLPLEVINGFQISPEARRVKFNT
LETLTHSSGPASTGRPRVAAA"
19700..20443
/notes="contains similarity to surface protein"
gene_id:MLN1.5"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB10980.1"
/db_xref="GI:9759514"
/translation="MATPTELLLLTAVFLSTTAQRAAPAGPAGPINITALEKG
GQFVTLRLNTTQIGNQINQINSSSEGTVLAPTNAQNKLKPGTLNKLSPDDQVK
LILYHVSQPFYTLDDLSVSNPRTQASGRDVGQVGLNTGQGNVNTGVVETRL
STLQRQPLAVYVDVLLPEEMFGRKISPMAPPKSPDVSDDESSESKAAAPS
ESKSGSGEMTGLGLGLVLCLELL"
join(21063..21237,21296..21408,21497..21667,21741..21953,
22037..22201)
/notes="gene_id:MLN1.6"
/codon_start=1
/evidence="not_experimental"
/product="prohibitin"
/protein_id="BAB10981.1"
/db_xref="GI:9759515"
/translation="MNVKVPNVPGSPALSALLKLVIGGLYCISSMNVNDGGRH
AIVFNRTGKDRVYPGTHFKPLPERALIYDVRSPYNSQTSGNSDQLQVTVIGR
VLTRPMGRDLPRIYTLGQNGERVLPSIINETLKAVAYQINASHLITQEVYRIGR
KIYTERAAKNALDDVSTINLFGKEFTAEIKKQVAAQEAERAKFIVEKAQDDKS
AIIRAQGEAKSAQLIGQAIANNEAFITLRKIEAREIAQTIAKSANKVYLNSSDLLIS
```

```
CDS
complement(join(22369..22845,22924..23036,23129..23230,
23307..23673))
/notes="gene_id:MLN1.7
unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB10982.1"
/db_xref="GI:9759516"
/translation="MDSKSLAKSKRAHTLHHSSKSHSVHKPKVPGVSEKNPEKLQGNQ
TKSPQSVRRVSALPNWDYDDELDAAEDSSILSHSDVIVPKSGADYLHLISEAQAE
NSKIENLDCLSLDDLLHDEFRRVSGMISARCEGILSHMEDDNFVYVEDGSGVKG
FLSLNVLAKTLENVDLHERLYIDPDLPLPELNTSQTKVRNEEPSSHIAQNDPI
VYQESSVREASLDQVKDILILDESEKSSAIEDLDDLNSSEAHQNPNVASAS
GKSAFETLSDLSKSHSSTQFNKPGNPQDKIHMTGFNDVLDLLESTVPSITPQS
NOTSKVLDLDFDSWLDTI"
join(31663..31801,32230..32626,33221..34085)
/notes="contains similarity to zinc finger protein ID1
gene_id:MLN1.8"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB10983.1"
/db_xref="GI:9759517"
/translation="MTSEVLQTISSGSGFAQPSSTLDHDESINPLVKKKRLPG
NPDPPEARVIALSPPTLMATNRFCEVCQGFQDQNLQHRHNLPMKLRKTSKEV
KRKYVPEKTCVHHSSRGALDITGKKHCFKRGHEKKTCEKAKRYAQVQSMKAH
SKTGYREYRCDCGTIFSRSDSFTHRAFCALAEETAKINAVSHLNLGAAGAPGV
NLNYOYLMGTPIPLQPVPOPTNPNNHHOHFOPPTSSLSLAWGQDIAEPQOPDY
DMVFNKAASACIDNNNTHDEQITONNASLTMTTTLASAFSLSSDQPPQANANSV
NWSATAQKAAEIGATSTTAANDPSTFLOSPLKSTDTQTTSYDSEKGFALFGSN
NNIGLSRSHDQEIENARNDTVASALDELQNYPMKRRRYDGGGEGGGGQTRDFLG
VGVTLCHPSSINGWI"
38966..39670
/notes="gene_id:MLN1.9
pir||T38261
similar to unknown protein"
/codon_start=1
/evidence="not_experimental"
/protein_id="BAB10984.1"
/db_xref="GI:9759518"
/translation="MKFTDSVIDLTVNGTKLSTQDNGSMHVGTSVWPCSILLSKFA
ERWTLSSSTTSPNPYAEFLFERRRGIELGTCGVAGMAFYLLGLTEIVLTDIAP
VMPALKNLKKNKTKALGSKLSTSVYNNRQDQISALPKPFDLVTAAVYVIEESYQGL
VTAMELVADDDQVLLGYQIRSEADKLFWELCDIVFKIEKVPHEHLHSDYAEETDV
YIFRKVKKNEAESVS"
join(41744..41855,41947..42360,42441..42520,42621..43001,
43096..43185,43267..43431,43518..43740,43847..43947,
44084..44170,44251..44403,44498..44653,44728..44937,
45026..45676,45764..46294,46387..46947,47129..47233,
47233..47233)
Query Match 85.5%; Score 18.8; DB 8; Length 84544;
Best Local Similarity 90.9%; Pred. No. 9.2;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| |||||
Db 16719 AACGAGTCCAGCTCGACCAGC 16698

RESULT 13
AC007864
LOCUS AC007864 167645 bp DNA HTG 17-JUL-2001
DEFINITION Trypanosoma brucei chromosome II clone RPC193-28H13, *** SEQUENCING
IN PROGRESS ***, 1 ordered pieces.
ACCESSION AC007864
VERSION AC007864.4 GI:14787195
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 167645)
AUTHORS El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K.,
```



Db 918 GAGTGTCTCAGCTAGACCAG 935

Search completed: February 25, 2002, 18:03:03  
Job time: 18591 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:41 ; Search time 716.55 Seconds  
(without alignments)  
26.322 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgctagctagaccagc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	AAH25457	PCR primer for end
2	22	100.0	22	AAH25457	PCR primer B03, to
3	22	100.0	3113	AAH25457	Cruciferin A gene.
C 4	17.2	78.2	20633	AAH25457	Enterococcus faeca
C 5	16.8	76.4	1692	AAH25457	NLERK2 cDNA 3' reg
6	16.2	73.6	353	AAH25457	Pseudomonas aerugi
7	16.2	73.6	353	AAH25457	Pseudomonas aerugi
8	15.8	71.8	716	AAH25457	Human cDNA clone (
9	15.8	71.8	850	AAH25457	Human cDNA clone (
10	15.8	71.8	1790	AAH25457	Human cDNA sequenc
11	15.8	71.8	1929	AAH25457	Human cDNA sequenc

12	15.8	71.8	2096	22	AAH25457	Human secreted pro
C 13	15.8	71.8	10708	19	AAH25457	Sequence of mouse
14	15.6	70.9	163	22	AAH25457	Probe #21465 used
15	15.6	70.9	578	22	AAH25457	Probe #8384 used t
C 16	15.6	70.9	695	21	AAH25457	Aspergillus oryzae
C 17	15.6	70.9	1193	21	AAH25457	Arabidopsis thalia
C 18	15.6	70.9	1236	21	AAH25457	Arabidopsis thalia
C 19	15.6	70.9	1238	21	AAH25457	Arabidopsis thalia
C 20	15.6	70.9	1239	21	AAH25457	Arabidopsis thalia
C 21	15.6	70.9	1240	21	AAH25457	Arabidopsis thalia
C 22	15.6	70.9	1968	21	AAH25457	Human secreted pro
23	15.6	70.9	2105	22	AAH25457	Human polynucleoti
C 24	15.6	70.9	2224	22	AAH25457	Human polynucleoti
C 25	15.6	70.9	3192	21	AAH25457	Arabidopsis thalia
C 26	15.6	70.9	8390	21	AAH25457	BAC containing rep
C 27	15.6	70.9	90336	21	AAH25457	BAC containing rep
C 28	15.4	70.0	307	21	AAH25457	Human secreted exp
C 29	15.4	70.0	1256	21	AAH25457	DNA encoding a kap
C 30	15.4	70.0	1330	16	AAH25457	Mouse oploid recep
C 31	15.4	70.0	2518	21	AAH25457	DNA encoding a kap
C 32	15.4	70.0	2600	16	AAH25457	Mouse kappa-3 opio
C 33	15.4	70.0	2634	21	AAH25457	DNA encoding a kap
C 34	15.4	70.0	4315	20	AAH25457	Enterococcus faeca
C 35	15.2	69.1	191	22	AAH25457	Probe #23963 used
C 36	15.2	69.1	350	13	AAH25457	MAD-3. Homo sapie
37	15.2	69.1	520	22	AAH25457	Probe #10878 used
38	15.2	69.1	657	15	AAH25457	Rat proteosome RIN
39	15.2	69.1	754	22	AAH25457	Drosophila melanog
C 40	15.2	69.1	819	18	AAH25457	Human inhibitory k
41	15.2	69.1	930	20	AAH25457	Human SOCS14 cDNA
C 42	15.2	69.1	1148	22	AAH25457	Pseudomonas sp Typ
C 43	15.2	69.1	1548	19	AAH25457	Human IkappaB poly
C 44	15.2	69.1	1550	13	AAH25457	Encodes IkB NF-kap
C 45	15.2	69.1	1550	19	AAH25457	Human I-kappa-B-al

ALIGNMENTS

RESULT 1

AAH25457

ID AAH25457 standard; DNA; 22 BP.

AC AAH25457;

DT 22-AUG-2001 (first entry)

XX PCR primer for endogenous sequences in transgenic plants.

DE Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
KW fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.

XX Synthetic.

XX WO200141538-A1.

XX 14-JUN-2001.

XX 06-DEC-2000; 2000WO-EP12872.

XX 08-DEC-1999; 99US-0457037.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX De Both G, De Beuckeleer M;

XX WPI; 2001-381419/40.

PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
PT with improved qualities, comprises a male-sterility gene and fertility  
PT restorer gene, integrated into the genome -  
XX Example 5; Page 53; 98pp; English.

XX The specification describes a pair of transgenic winter oilseed rape  
CC plants suited for producing hybrid seed. One of the plants has an  
CC expression cassette comprising a male-sterility gene (e.g. barnase  
CC gene), and the other plant has an expression cassette comprising a  
CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
CC The fertility restorer gene is capable of preventing the activity of the  
CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
CC Plants developed from the hybrid seed have agronomic performance,  
CC genetic stability and adaptability to different genetic backgrounds.  
CC PCR primers AA25457-58 were used to amplify endogenous sequences  
CC from transgenic plants of the invention.  
XX  
SQ Sequence 22 BP; 7 A; 6 C; 6 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.043; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
|||||  
Db 1 aacgagtgctcagctagaccagc 22

RESULT 2  
AA07002  
ID AAD07002 standard; DNA; 22 BP.  
XX  
AC AAD07002;  
XX  
XX  
DT 06-AUG-2001 (first entry)  
DE PCR primer B03, to recognise foreign DNA and flanking sequence of MS-B2.  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; PCR primer; ss.  
XX  
XX Brassica napus.  
XX  
XX WO200131042-A2.  
XX  
XX 03-MAY-2001.  
XX  
XX 26-OCT-2000; 2000WO-EP10680.  
XX  
XX 29-OCT-1999; 99US-0430497.  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
XX Weston B, De Beuckeleer M;  
XX  
XX WPI; 2001-300517/31.  
XX  
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX  
XX Example 5; Page 33; 53pp; English.

CC The present invention relates to a transgenic Brassica plant or its  
CC seed, cells or tissues, characterised by harbouring a specific  
CC transformation event, particularly by the presence of a male-sterility  
CC gene, at a specific location in the Brassica genome. Transgenic  
CC Brassica plant is useful for producing a hybrid seed by crossing the  
CC transgenic plant with a male-fertile Brassica plant and harvesting the  
CC hybrid seed from the transgenic Brassica plant.  
CC The present sequence is PCR primer which is used to recognise foreign  
CC DNA and a flanking sequence of elite event MS-B2.  
XX  
XX Sequence 22 BP; 7 A; 6 C; 6 G; 3 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.043; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
|||||  
Db 1 aacgagtgctcagctagaccagc 22

RESULT 3  
AAQ13870  
ID AAQ13870 standard; DNA; 3113 BP.  
XX  
AC AAQ13870;  
XX  
XX  
DT 09-DEC-1991 (first entry)  
DE Cruciferin A gene.  
XX  
XX Seed storage protein; crua; ss.  
XX  
XX Brassica napus.  
XX  
XX EP449376-A.  
XX  
XX 02-OCT-1991.  
XX  
XX 25-MAR-1991; 91EP-0200688.  
XX  
XX 25-MAR-1991; 91EP-0200688.  
XX 23-MAR-1990; 90US-0498561.  
XX  
XX (KONN ) GIST-BROCADES NV.  
XX  
XX Pen J, Stijmons PC, Van Ooyen AJJ, Rietveld K, Verwoerd T;  
PI Quax WJ;  
XX  
XX WPI; 1991-289815/40.  
XX  
XX Seeds contg. enhanced enzyme levels from transgenic plants - used  
PT for catalysing reactions, increasing nutritional values or  
PT treating digestive disorders.  
XX  
XX Example; Fig 3; 38pp; English.

CC The DNA is the genomic sequence of the seed storage protein gene  
CC cruciferin A (crua). It can be used in the prodn. of transgenic  
CC plants expressing cruciferin in its seeds for use in an industrial  
CC process. The seeds contg. the cruciferin can be used without the  
CC need for first extracting and/or isolating the enzymes. The use of  
CC seeds for the storage of cruciferin provides a stable vehicle which  
CC is easily packaged and transported and easily handled during use.  
XX  
XX See also AAQ13871-Q13877.  
XX  
XX Sequence 3113 BP; 961 A; 685 C; 586 G; 881 T; 0 other;

Query Match 100.0%; Score 22; DB 12; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.09; Mismatches 0; Indels 0; Gaps 0;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgctcagctagaccagc 22  
|||||  
Db 790 aacgagtgctcagctagaccagc 811

RESULT 4  
AAQ13213/c  
ID AAX13213 standard; DNA; 20633 BP.  
XX  
AC AAX13213;  
XX  
XX 19-MAR-1999 (first entry)

```
XX DE Enterococcus faecalis genome contig SEQ ID NO:276.
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
XX KW vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
XX PN WO9850555-A2.
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Barash SC, Dillon PJ, Kunsch CA;
XX PI WPI: 1999-045171/04.
XX DR
XX XX New isolated Enterococcus faecalis polynucleotides and polypeptides
XX PT - used to develop products for the detection of Enterococcus and for
XX PT use in vaccines for prevention or attenuation of Enterococcus
XX PT infection.
XX XX Claim 1; Page 1284-1294; 2084pp; English.
XX PS
XX CC A computer readable medium has been developed which has recorded on it
XX CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX CC AAX12938 to AAX13919 represent these nucleotide sequences which are
XX CC primary nucleotide sequences, also known as contigs. The computer-based
XX CC system can identify fragments of the Enterococcus faecalis genome with
XX CC commercial importance. The products can be used to detect the presence
XX CC of Enterococcus faecalis in samples. They can also be used for
XX CC diagnosing Enterococcal infection in an animal and monitoring
XX CC progression of disease, and for identifying agents which can be used to
XX CC modulate the growth or pathogenicity of Enterococcus faecalis, or
XX CC another related organism, in vivo or in vitro. In particular the
XX CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX CC can be used in vaccines to prevent or attenuate an Enterococcal
XX CC infection.
XX SQ Sequence 20633 BP; 5857 A; 4284 C; 3025 G; 7427 T; 40 other;

Query Match 78.2%; Score 17.2; DB 20; Length 20633;
Best Local Similarity 86.4%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22
||||| ||| ||| |||||
Db 18911 AACGAGTATCATCTAAACCAGC 18890

RESULT 5
AAT60974/c
ID AAT60974 standard; cDNA; 1692 BP.
XX AC
XX AC AAT60974;
XX XX
XX DT 23-JUN-1997 (first entry)
XX DE
XX DE NLERK2 cDNA 3' region.
XX KW LERK; ligand for eph-related kinase; ERK; NLERK2;
XX KW receptor protein tyrosine kinase; cell proliferation;
XX KW cell differentiation; cell survival; nerve cell; ss.
XX OS Homo sapiens.
```

```
XX PN WO9704091-A1.
XX PD 06-FEB-1997.
XX PF 19-JUL-1996; 96WO-AU00460.
XX PR 05-FEB-1996; 96AU-0007890.
XX PR 20-JUL-1995; 95AU-0004263.
XX PR 27-NOV-1995; 95AU-0006847.
XX PR 22-DEC-1995; 95AU-0007299.
XX PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX PI Nicola NA;
XX DR WPI: 1997-132632/12.
XX XX Nucleic acid mol. encoding ligand for eph-related kinase - useful
XX PT for treatment of, pref. neuronal, cells to increase survival,
XX PT proliferation and differentiation
XX XX Claim 5; Page 43-45; 71pp; English.
XX CC The 3' region (AAT60974) and 5' region (AAT60975) sequences were
XX CC detd. for a cDNA sequence (see also AAT60966) coding for NLERK2
XX CC (AAW10637), a novel human ligand for eph-related kinase (LERK). The
XX CC NLERK2 cDNA was obt'd. from a human foetal brain cDNA library using
XX CC probes (see also AAT60967-69) based on an expressed sequence tag
XX CC previously isolated on the basis of homology to conserved regions
XX CC of known LERKs (see also AAW10633-36).
XX SQ Sequence 1692 BP; 316 A; 504 C; 385 G; 472 T; 15 other;

Query Match 76.4%; Score 16.8; DB 18; Length 1692;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acgagtgtcagctagaccagc 21
||||| ||||| ||||| |||||
Db 1434 ACCACTGTGCTAGACCAGC 1415

RESULT 6
AAF15388
ID AAF15388 standard; DNA; 353 BP.
XX AC
XX AC AAF15388;
XX XX
XX DT 09-MAR-2001 (first entry)
XX DE
XX DE Pseudomonas aeruginosa ssrA gene, SEQ ID NO:101.
XX KW ssrA gene; tmRNA; bacterium; chloroplast; diatom; detection;
XX KW detection; identification; quantification; characterisation;
XX KW nucleic acid array; DNA chip; drug design; treatment monitoring;
XX KW contamination; ds.
XX OS Pseudomonas aeruginosa.
XX PN WO200070086-A1.
XX XX
XX PD 23-NOV-2000.
XX PF 15-MAY-2000; 2000WO-IE000066.
XX PR 14-MAY-1999; 99WO-IE000043.
XX XX (IRBI-) ENTERPRISE IRELAND T/A BIOTRESEARCH IRELA.
XX PA (UYNA-) UNIV NAT IRELAND GALWAY.
XX XX
XX PI Barry TG, Smith TJ;
```

XX WPI; 2001-025025/03.

XX Use of ssrA gene, tmRNA, or fragments of them, as target regions in

XX probe assays for detection of prokaryotic or eukaryotic organisms, and

XX for determination of species -

XX Disclosure; Page 55; 221pp; English.

XX The invention relates to the ssrA gene or tmRNA, an RNA transcript

XX of the ssrA gene, or fragments thereof as target regions in a nucleic

XX acid probe assay for the detection and identification of prokaryotic

XX and/or eukaryotic organisms. The invention also relates to 38 novel ssrA

XX sequences and their tmRNA transcripts (claimed), and to pan-bacterial,

XX genus- and species-specific ssrA gene/tmRNA-directed probes and PCR

XX primers (claimed). tmRNA is a stable, high copy number RNA which is

XX found in all bacteria and is also found in chloroplasts and diatoms. It

XX has a dual function both as a tRNA and as an mRNA and is involved in

XX rescuing truncated mRNAs which have lost stop codons. ssrA genes and

XX tmRNA sequences can be used as target regions in nucleic acid probe

XX assays for the detection, identification, or quantification of a

XX prokaryotic or eukaryotic organism. cDNA transcripts of tmRNA molecules

XX may also be used as probes for in vitro or in situ nucleic acid

XX hybridisation assays. A fragment of the ssrA gene or a tmRNA molecule

XX corresponding to a region of high homology from the 5' end or the 3'

XX end of the DNA molecule can be used as a universal target region in

XX a nucleic acid probe assay, while a fragment of the ssrA gene or a

XX tmRNA molecule corresponding to a region of low homology can be used

XX as a target region to distinguish between species and as a target

XX region for the generation of genus-specific probes. These regions

XX may also be used as the basis for amplification primer design. The

XX target regions may be used as the basis of an assay for distinguishing

XX between living and dead prokaryotic or eukaryotic organisms, and in

XX a multiple probe format for broad scale detection and/or identification

XX of prokaryotic or eukaryotic organisms. An ssrA gene probe or a

XX tmRNA transcript probe can be linked to a microarray gene chip system

XX for the broad scale high throughput detection and identification of

XX prokaryotic or eukaryotic organisms. A fragment of the ssrA gene or

XX the tmRNA transcript can be used in an assay to obtain a DNA profile

XX of a prokaryotic organism and distinguish between strains of the same

XX species. The ssrA gene, the tmRNA transcript, DNA complementary to

XX an ssrA gene or tmRNA, or a fragment thereof can be used to design an

XX agent directed against infectious prokaryotic or eukaryotic organisms

XX for therapeutic purposes, and target regions may be used to monitor the

XX efficacy of drug therapies against infectious agents. Target regions may

XX also be used to monitor the viability and level of probiotic organisms

XX in the gastrointestinal tract. The methods and nucleic acids and

XX compositions of the invention have applications in medicine, and also

XX in industry (e.g., for assessing bacterial contamination of a foodstuff

XX or an environmental sample). Sequences AAF15338-F15442 represent ssrA

XX genes, or fragments thereof, from a wide variety of organisms.

XX Sequence 353 BP; 94 A; 94 C; 99 G; 66 T; 0 other;

Query Match 73.6%; Score 16.2; DB 22; Length 353;

Best Local Similarity 85.7%; Pred. No. 57;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagacag 21

||||| ||||| ||||| |||||

Db 166 aacgagtgtcagctagacag 186

RESULT 7

AAF15493

ID AAF15493 standard; RNA; 353 BP.

AC AAF15493;

XX

XX 09-MAR-2001 (first entry)

XX Pseudomonas aeruginosa tmRNA, SEQ ID NO:102.

DE

ssrA gene; tmRNA; bacterium; chloroplast; diatom; detection;

detection; identification; quantification; characterisation;

nucleic acid array; DNA chip; drug design; treatment monitoring;

contamination; ss.

Pseudomonas aeruginosa.

WO200070086-A1.

23-NOV-2000.

15-MAY-2000; 2000WO-IE00066.

14-MAY-1999; 99WO-IE00043.

(IRBI-) ENTERPRISE IRELAND T/A BIORESEARCH IRELA.

(UYNNA-) UNIV NAT IRELAND GALWAY.

Barry TG, Smith TJ;

WPI; 2001-025025/03.

Use of ssrA gene, tmRNA, or fragments of them, as target regions in

probe assays for detection of prokaryotic or eukaryotic organisms, and

for determination of species -

Disclosure; Page 55; 221pp; English.

The invention relates to the ssrA gene or tmRNA, an RNA transcript

of the ssrA gene, or fragments thereof as target regions in a nucleic

acid probe assay for the detection and identification of prokaryotic

and/or eukaryotic organisms. The invention also relates to 38 novel ssrA

sequences and their tmRNA transcripts (claimed), and to pan-bacterial,

genus- and species-specific ssrA gene/tmRNA-directed probes and PCR

primers (claimed). tmRNA is a stable, high copy number RNA which is

found in all bacteria and is also found in chloroplasts and diatoms. It

has a dual function both as a tRNA and as an mRNA and is involved in

rescuing truncated mRNAs which have lost stop codons. ssrA genes and

tmRNA sequences can be used as target regions in nucleic acid probe

assays for the detection, identification, or quantification of a

prokaryotic or eukaryotic organism. cDNA transcripts of tmRNA molecules

may also be used as probes for in vitro or in situ nucleic acid

hybridisation assays. A fragment of the ssrA gene or a tmRNA molecule

corresponding to a region of high homology from the 5' end or the 3'

end of the DNA molecule can be used as a universal target region in

a nucleic acid probe assay, while a fragment of the ssrA gene or a

tmRNA molecule corresponding to a region of low homology can be used

as a target region to distinguish between species and as a target

region for the generation of genus-specific probes. These regions

may also be used as the basis for amplification primer design. The

target regions may be used as the basis of an assay for distinguishing

between living and dead prokaryotic or eukaryotic organisms, and in

a multiple probe format for broad scale detection and/or identification

of prokaryotic or eukaryotic organisms. An ssrA gene probe or a

tmRNA transcript probe can be linked to a microarray gene chip system

for the broad scale high throughput detection and identification of

prokaryotic or eukaryotic organisms. A fragment of the ssrA gene or

the tmRNA transcript can be used in an assay to obtain a DNA profile

of a prokaryotic organism and distinguish between strains of the same

species. The ssrA gene, the tmRNA transcript, DNA complementary to

an ssrA gene or tmRNA, or a fragment thereof can be used to design an

agent directed against infectious prokaryotic or eukaryotic organisms

for therapeutic purposes, and target regions may be used to monitor the

efficacy of drug therapies against infectious agents. Target regions may

also be used to monitor the viability and level of probiotic organisms

in the gastrointestinal tract. The methods and nucleic acids and

compositions of the invention have applications in medicine, and also

in industry (e.g., for assessing bacterial contamination of a foodstuff

or an environmental sample). Sequences AAF15338-F15442 represent ssrA

genes, or fragments thereof, from a wide variety of organisms.

Sequence 353 BP; 94 A; 94 C; 99 G; 66 T; 0 other;

```

Query Match      73.6%; Score 16.2; DB 22; Length 353;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 15; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 1 aacgagtgctcagctagaccag 21
    |||||:|||||:|||||
Db 166 aacgacugacagauagaacag 186

RESULT 8
AAH05632
ID AAH05632 standard; cDNA; 716 BP.
XX AC AAH05632;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:2467.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX FN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PR Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 1; SEQ ID 2467; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesising 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesising polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.

Query Match      71.8%; Score 15.8; DB 22; Length 716;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 4 gagtgctcagctagaccagc 22
    ||||| ||||| |||||
Db 122 gagtggcagctggaccagc 140

RESULT 9
AAH07552
ID AAH07552 standard; cDNA; 850 BP.
XX AC AAH07552;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:4387.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX FN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PR Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 1; SEQ ID 4387; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesising 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesising polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
```

CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 850 BP; 196 A; 227 C; 258 G; 160 T; 9 other;

Query Match 71.8%; Score 15.8; DB 22; Length 850;  
 Best Local Similarity 89.5%; Pred. No. 1e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gaggtcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 126 gagtggcagctggaccagc 144

RESULT 10  
 AAH14119  
 ID AAH14119 standard; cDNA; 1790 BP.

AC AAH14119;  
 XX  
 XX  
 XX 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:11310.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX OS

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX XX

XX XX 28-JUL-2000; 2000EP-0116126.

XX XX

XX XX 29-JUL-1999; 99JP-0248036.

XX XX 27-AUG-1999; 99JP-0300253.

XX XX 11-JAN-2000; 2000JP-0118776.

XX XX 02-MAY-2000; 2000JP-0183767.

XX XX 09-JUN-2000; 2000JP-0241899.

XX XX (HELI-) HELIX RES INST.

XX XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX SQ Sequence 1790 BP; 398 A; 456 C; 504 G; 432 T; 0 other;

Query Match 71.8%; Score 15.8; DB 22; Length 1790;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gagtgcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 122 gagtggcagctggaccagc 140

RESULT 11  
 AAH16554  
 ID AAH16554 standard; cDNA; 1929 BP.

XX AC AAH16554;  
 XX  
 XX 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:15615.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX XX

XX XX 28-JUL-2000; 2000EP-0116126.

XX XX

XX XX 29-JUL-1999; 99JP-0248036.

XX XX 27-AUG-1999; 99JP-0300253.

XX XX 11-JAN-2000; 2000JP-0118776.

XX XX 02-MAY-2000; 2000JP-0183767.

XX XX 09-JUN-2000; 2000JP-0241899.

XX XX (HELI-) HELIX RES INST.

XX XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX SQ Sequence 1929 BP; 438 A; 489 C; 536 G; 466 T; 0 other;

Query Match 71.8%; Score 15.8; DB 22; Length 1929;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gagtgcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 126 gagtgcagctgagaccagc 144

RESULT 12  
 AAD05091  
 ID AAD05091 standard; cDNA; 2096 BP.

XX AC AAD05091;

XX DT 17-JUL-2001 (first entry)

XX DE Human secreted protein-encoding gene 8 cDNA clone HSYAZ50, SEQ ID NO:49.

XX KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angioinetic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnary; gene therapy;  
 KW cell culture; chemotaxis; food additive;  
 KW binding partner identification; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 XX CDS 2..841

XX FT /\*tag= a  
 XX FT /product= "Human secreted protein"  
 XX FT /note= "CDS does not include start codon"

XX FT /partial

XX FT sig\_peptide 2..4

XX FT /\*tag= b

XX FT mat\_peptide 5..838

XX FT /\*tag= c  
 XX FT /product= "Mature human secreted protein"

XX WO200134768-A2.

XX PN 17-MAY-2001.

XX PD 01-NOV-2000; 2000WO-US30039.

XX PR 09-NOV-1999; 99US-0164344.

XX PR 07-APR-2000; 2000US-0195296.

XX PR 27-JUL-2000; 2000US-0221367.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Olsen HS, Komatsoulis G, Duan DR, Ebner R, Ruben SM;

XX DR WPI; 2001-308780/32.

XX DR P-PSDB; AAE01202.

XX PT Isolated nucleic acid molecule encoding a human secreted protein is

PT used in preventing, treating or ameliorating a medical condition -  
 XX Claim 1; Page 429; 474pp; English.

XX CC AAD05053-AAD05106 represent cDNAs corresponding to 15 human secreted  
 CC protein genes, and AAE01164-AAE01217 represent the proteins they encode.  
 CC AAE01218-AAE01226 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the  
 CC amount of the new protein in a sample or by determining the presence of  
 CC mutations in the new genes. Specific uses are described for each of the  
 CC 15 genes, based on the tissues in which they are most highly expressed,  
 CC and include developing products for the diagnosis or treatment of  
 CC proliferative disorders, cancer, tumours, foetal and developmental  
 CC abnormalities, haematopoietic disorders, diseases of the immune system,  
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
 CC allergies, neurological disorders (e.g., Alzheimer's disease,  
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
 CC cardiovascular disorders, angioinetic disorders, kidney disorders,  
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine  
 CC disorders, and infections. The proteins can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues, to identify their  
 CC cognate ligands or binding partners, and in chemotaxis, and can be used  
 CC as a food additive or preservative to modify storage properties.  
 CC Antibodies specific for a protein of the invention can be used in  
 CC alleviating symptoms associated with the disorders mentioned above, and  
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
 CC immunosorbent assay (ELISA). The present sequence represents a human  
 CC secreted protein-encoding cDNA of the invention.

XX SQ Sequence 2096 BP; 476 A; 529 C; 594 G; 504 T; 3 other;

Query Match 71.8%; Score 15.8; DB 22; Length 2096;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 gagtgcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 149 gagtgcagctgagaccagc 167

RESULT 13

AAV69286/c

ID AAV69286 standard; DNA; 10708 BP.

XX AC AAV69286;

XX DT 02-FEB-1999 (first entry)

XX DE Sequence of mouse activin genetic loci.

XX KW Liver activin; beta c; beta e; cell differentiation; haematopoiesis;  
 KW erythroid; ovarian follicular maturation; hormone; neuronal survival;  
 KW spermatogenesis; bone; insulin; cardiac morphogenesis; osteoporosis;  
 KW osteomalacia; erythropoiesis; haemophilia; cystic fibrosis; immunoassay;  
 KW menstrual disorder; transgenic; modulator; ss.

XX OS Mus sp.

XX FH Key Location/Qualifiers  
 XX CDS 1147..2736

XX FT /\*tag= a  
 XX FT /product= activin beta c"  
 XX FT /note= "contains introns"

XX FT exon 1147..1462

XX FT /\*tag= b

XX FT /number= 1  
 XX FT /note= "activin beta c exon 1"

FT intron 1463..1994  
 FT /\*tag= c  
 FT /number= 1  
 FT /note= "activin beta c partial intron (12-kb)"  
 FT exon 1995..2733  
 FT /\*tag= d  
 FT /number= 2  
 FT 2737..8248  
 FT promoter  
 FT /\*tag= e  
 FT /note= "activin beta e promoter region"  
 FT CDS 8249..9536  
 FT /\*tag= f  
 FT /product= "activin beta e"  
 FT /note= "contains introns"  
 FT exon 8249..8546  
 FT /\*tag= g  
 FT /number= 1  
 FT /note= "activin beta e exon 1"  
 FT intron 8547..8782  
 FT /\*tag= h  
 FT /number= 1  
 FT /note= "activin beta e intron"  
 FT exon 8783..9533  
 FT /\*tag= i  
 FT /number= 2  
 FT /note= "activin beta e exon 2"  
 XX  
 PN W09822492-A1.  
 XX  
 XX 28-MAY-1998.  
 XX  
 XX 20-NOV-1997; 97WO-US20882.  
 XX  
 XX 20-NOV-1996; 96US-0752919.  
 XX  
 XX (UNMI ) UNIV MICHIGAN.  
 XX  
 XX Bonadio J, Fang J;  
 XX WPI; 1998-312408/27.  
 DR P-PSDB; AAW60617, AAW60618.  
 XX  
 XX New isolated nucleic acid encoding sub-units of liver activin -  
 PT useful for regulating growth and differentiation of cells, e.g. for  
 PT treating liver, bone and haematopoietic disorders  
 XX  
 PS Disclosure; Fig 4D-G; 141pp; English.  
 XX  
 CC This represents the sequence of mouse activin genetic loci. The invention  
 CC relates to murine beta c and beta e polypeptides and the genes encoding  
 CC them. Disorders of cell growth or differentiation (or susceptibility to  
 CC them) are diagnosed by measuring liver activin gene activity or by  
 CC detecting a mutation in the liver activin gene. Disorders of  
 CC haematopoiesis, erythroid differentiation, ovarian follicular maturation,  
 CC insulin secretion, neuronal survival, spermatogenesis, bone formation,  
 CC are diagnosed using the liver activin. Cell growth and differentiation can  
 CC be stimulated by treatment with an liver activin compound or agent that  
 CC upregulates the compound's expression. Antagonists can be used to treat  
 CC liver diseases while agonists can be used to increase growth and  
 CC regeneration of liver tissue. The liver activin compound may also induce  
 CC bone growth (e.g. for treating osteoporosis or osteomalacia) or  
 CC haematopoiesis, particularly erythropoiesis, e.g. for treating  
 CC useful in immunoassays, to generate anti-idiotypic antibodies (which bind  
 CC to liver activin receptors) and to inhibit liver activin. Also,  
 CC transgenic animals containing liver activin gene can be used to produce  
 CC the liver activin (in correctly processed and modified forms) proteins,  
 CC or the transgenic animals, are useful for screening for liver activin  
 CC modulators.  
 CC This cDNA encodes a murine liver activin beta c polypeptide. Sequences  
 CC derived from beta c cDNA clone is used for screening and cloning a liver  
 CC activin beta e gene. Disorders of cell growth or differentiation (or

CC susceptibility to them) are diagnosed by measuring liver activin gene  
 CC activity or by detecting a mutation in the liver activin gene. Disorders  
 CC of haematopoiesis, erythroid differentiation, ovarian follicular  
 CC maturation, hormone secretion, neuronal survival, spermatogenesis, bone  
 CC formation, insulin secretion or cardiac morphogenesis are some conditions  
 CC that can be diagnosed using the liver activin. Cell growth and  
 CC differentiation can be stimulated using the liver activin. Antagonists  
 CC compound or agent that upregulates the compound's expression. Antagonists  
 CC can be used to treat liver diseases while agonists can be used to  
 CC increase growth and regeneration of liver tissue. The liver activin  
 CC compound may also induce bone growth (e.g. for treating osteoporosis or  
 CC osteomalacia) or haematopoiesis, particularly erythropoiesis, e.g. for  
 CC treating haemophilia, cystic fibrosis or menstrual disorders. Antibodies  
 CC are useful in immunoassays, to generate anti-idiotypic antibodies (which  
 CC bind to liver activin receptors) and to inhibit liver activin. Also,  
 CC transgenic animals containing liver activin gene can be used to produce  
 CC the liver activin (in correctly processed and modified forms) proteins,  
 CC or the transgenic animals, are useful for screening for liver activin  
 CC modulators.  
 XX

SQ Sequence 10708 BP; 2651 A; 2844 C; 2607 G; 2546 T; 60 other;

Query Match 71.8%; Score 15.8; DB 19; Length 10708;  
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gagtgcagctagaccagc 22  
 ||||| |||||  
 Db 1949 GAGTGTCTGTCGACCAGC 1931

## RESULT 14

AA152779  
 ID AA152779 standard; DNA; 163 BP.

XX  
 AC AA152779;

XX 17-OCT-2001 (first entry)

DE Probe #21465 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -

XX Claim 25; SEQ ID No 21465; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SEN).



**Qy**      1   aacgagtgtcagctagaccagc 22  
         ||||| ||| |||||||||  
**Db**     257 aaqgaqtctctgcctagaccagc 278

\_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:52 ; Search time 301.6 Seconds  
(without alignments)  
16.520 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgcagtagaccagc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/1na/PTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/1na/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	3113	1	US-08-146-422-20
2	22	100.0	3113	1	US-08-626-554-2
c 3	15.4	70.0	1330	3	US-08-147-592A-5
c 4	15.4	70.0	2600	1	US-08-147-949A-1
c 5	15.2	69.1	350	2	US-08-466-337A-16
c 6	15.2	69.1	350	2	US-08-475-359-16
c 7	15.2	69.1	350	3	US-08-465-887A-16
c 8	15.2	69.1	819	4	US-09-349-627-2
c 9	15.2	69.1	1550	2	US-08-466-337A-17
c 10	15.2	69.1	1550	2	US-08-475-359-17
c 11	15.2	69.1	1550	2	US-08-802-322-2
c 12	15.2	69.1	1550	3	US-08-465-887A-17
c 13	15.2	69.1	1550	3	US-08-895-601-3
c 14	14.6	66.4	2028	3	US-08-933-750C-73
c 15	14.6	66.4	2028	3	US-09-234-613-73
c 16	14.6	66.4	2915	2	US-08-500-857A-1
c 17	14.6	66.4	3719	1	US-08-920-812-10
c 18	14.6	66.4	3719	1	US-08-920-827-10
c 19	14.6	66.4	3719	1	US-08-921-177-10
c 20	14.6	66.4	3719	1	US-08-362-577C-10
c 21	14.6	66.4	3719	2	US-08-920-828-10
c 22	14.6	66.4	50341	1	US-08-247-901C-1
c 23	14.6	66.4	50341	2	US-09-075-904-1
c 24	14.6	66.4	52297	4	US-09-426-436-1
c 25	14.6	66.4	52297	4	US-08-705-557-1
c 26	14.6	66.4	4411529	4	US-09-103-840A-1
c 27	14.4	65.5	1458	5	PCT-US94-03437-1

c 28	14.4	65.5	1461	1	US-07-596-867C-10	Sequence 10, Appl
c 29	14.4	65.5	1506	3	US-08-445-463B-1	Sequence 1, Appl
c 30	14.4	65.5	1506	3	US-08-445-464C-1	Sequence 1, Appl
c 31	14.4	65.5	1963	5	PCT-US91-07715A-1	Sequence 1, Appl
c 32	14.4	65.5	1970	1	US-07-596-867C-1	Sequence 1, Appl
c 33	14.2	64.5	192	4	US-09-060-756-634	Sequence 634, App
c 34	14.2	64.5	215	4	US-09-060-756-194	Sequence 194, App
c 35	14.2	64.5	230	4	US-09-060-756-422	Sequence 422, App
c 36	14.2	64.5	579	2	US-08-611-757-50	Sequence 50, Appl
c 37	14.2	64.5	579	5	PCT-US95-05980-50	Sequence 50, Appl
c 38	14.2	64.5	1679	6	5196194-12	Patent No. 5196194
c 39	14.2	64.5	2389	2	US-08-372-652-8	Sequence 8, Appl
c 40	14.2	64.5	2389	5	PCT-US95-16311-8	Sequence 8, Appl
c 41	14.2	64.5	2940	3	US-09-195-868-11	Sequence 11, Appl
c 42	14.2	64.5	3262	2	US-08-678-039A-41	Sequence 41, Appl
c 43	14.2	64.5	3691	3	US-09-195-868-12	Sequence 12, Appl
c 44	14.2	64.5	4147	3	US-08-560-005-1	Sequence 1, Appl
c 45	14.2	64.5	4147	4	US-09-418-540-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-146-422-20  
; Sequence 20, Application US/08146422  
; Patent No. 5543576  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOLJEN, ALBERT J. J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERMOERD, TEUNIS C.  
; APPLICANT: QUAX, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/146.422  
; FILING DATE: 02-NOV-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNEDY, BILL  
; REGISTRATION NUMBER: 33,407  
; REFERENCE/DOCKET NUMBER: 44615-20011.23  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-146-422-20

Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
 Db 790 AACGAGTGTCTAGACCAGC 811

## RESULT 2

US-08-626-554-2  
 ; Sequence 2, Application US/08626554  
 ; Patent No. 5714474  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN OOIJEN, ALBERT J.J.  
 ; APPLICANT: RIETVELD, KRIJN  
 ; APPLICANT: HOEKEMA, ANDREAS  
 ; APPLICANT: PEN, JAN  
 ; APPLICANT: SIJMONS, PETER C.  
 ; APPLICANT: VERWOERD, TEUNIS C.  
 ; APPLICANT: QUAX, WILHEMUS J.  
 ; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
 ; TITLE OF INVENTION: USE  
 ; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 2000 PENNSYLVANIA AVENUE NW  
 ; CITY: WASHINGTON  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1888  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/626,554  
 ; FILING DATE: 02-APR-1996  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 26192-20011.10  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; TELEX: 90-4030 MRSNFOERSWSH  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3113 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-626-554-2

Query Match 100.0%; Score 22; DB 1; Length 3113;  
 Best Local Similarity 100.0%; Pred. No. 0.042;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
 Db 790 AACGAGTGTCTAGACCAGC 811

## RESULT 3

US-08-147-592A-5/c  
 ; Sequence 5, Application US/08147592A  
 ; Patent No. 6096513  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bell, Graeme I  
 ; APPLICANT: Relsine, Terry  
 ; APPLICANT: Yasuda, Kazuki  
 ; TITLE OF INVENTION: Opioid Receptor Genes,

; TITLE OF INVENTION: Compositions and Methods  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold, White & Durkee  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: Texas  
 ; COUNTRY: United States of America  
 ; ZIP: 77210

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/147,592A  
 ; FILING DATE: 05-NOV-1993  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/100,694  
 ; FILING DATE: 30-JUL-1993  
 ; CLASSIFICATION: 435  
 ; APPLICATION NUMBER: 08/066,296  
 ; FILING DATE: 20-MAY-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Wilson, Mark B.  
 ; REGISTRATION NUMBER: 37,259  
 ; REFERENCE/DOCKET NUMBER: ARCD:105  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512) 418-3000  
 ; TELEFAX: (512) 474-7577  
 ; TELEX: N/A  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1330 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: CDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 161..1261  
 ; US-08-147-592A-5

Query Match 70.0%; Score 15.4; DB 3; Length 1330;  
 Best Local Similarity 94.1%; Pred. No. 63;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 aatgtcagctagaccag 21  
 Db 462 AGTGTCTAGACCAG 446

## RESULT 4

US-08-147-949A-1/c  
 ; Sequence 1, Application US/08147949A  
 ; Patent No. 5747279  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pasternak, Gavrill W.  
 ; APPLICANT: Pan, Ying-Xian  
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
 ; TITLE OF INVENTION: KAPPA3 OPIOID RECEPTORS, RECEPTORS  
 ; TITLE OF INVENTION: ENCODED THEREBY, AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036

```

RESULT      5
US-08-466-337A-16/c
; Sequence 16, Application US/08466337A
; Patent No. 5830756
; GENERAL INFORMATION:
; APPLICANT: Haskill, John S.
; APPLICANT: Baldwin Jr., Albert S.
; APPLICANT: Ralph, Peter
; TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional
; TITLE OF INVENTION: Activator and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/ 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,337A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 0899.008/33518
; TELECOMMUNICATION INFORMATION:

```

Qy 3 cgagtgctagaccagc 22

4

TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 95...1045  
US-08-466-337A-17

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgtagcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGCTGGCCAGC 898

RESULT 10  
US-08-475-359-17/c  
Sequence 17, Application US/08475359  
Patent No. 5846714  
GENERAL INFORMATION:  
APPLICANT: Haskill, John S.  
APPLICANT: Baldwin Jr., Albert S.  
APPLICANT: Ralph, Peter  
TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional  
TITLE OF INVENTION: Activator and Uses Thereof  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower/ 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,359  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pochopien, Donald J.  
REGISTRATION NUMBER: 32,167  
REFERENCE/DOCKET NUMBER: 0899,004,33514  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 95...1045  
US-08-475-359-17

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgtagcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGCTGGCCAGC 898

RESULT 11  
US-08-802-322-2/c  
Sequence 2, Application US/08802322  
Patent No. 5932425  
GENERAL INFORMATION:  
APPLICANT: Alkalay, Irit  
APPLICANT: Ben-Neriah, Yinon  
APPLICANT: Ciechanover, Aaron  
APPLICANT: Manning, Anthony  
APPLICANT: Mercurio, Frank  
APPLICANT: Yaron, Avraham  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING  
TITLE OF INVENTION: CELLULAR NF-(B ACTIVATION  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,322  
FILING DATE: 18-FEB-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 860098.418  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-802-322-2

Query Match 69.1%; Score 15.2; DB 2; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgtagcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGCTGGCCAGC 898

RESULT 12  
US-08-465-887A-17/c  
Sequence 17, Application US/08465887A  
Patent No. 6001582  
GENERAL INFORMATION:  
APPLICANT: Haskill, John S.  
APPLICANT: Baldwin Jr., Albert S.  
APPLICANT: Ralph, Peter  
TITLE OF INVENTION: Inhibitor of NF-kB Transcriptional

;; TITLE OF INVENTION: Activator and Uses Thereof  
;; NUMBER OF SEQUENCES: 18  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
;; STREET: 6300 Sears Tower/ 233 South Wacker Drive  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: United States  
;; ZIP: 60606-6402  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/465,887A  
;; FILING DATE: 06-JUN-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fochoplen, Donald J.  
;; REGISTRATION NUMBER: 32,167  
;; REFERENCE/DOCKET NUMBER: 0899.006/33516  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312/474-6300  
;; TELEFAX: 312/474-0448  
;; TELEX: 25-3856  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1550 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 95..1045  
US-08-465-887A-17

Query Match 69.1%; Score 15.2; DB 3; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgctcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGCTGCCGAGC 898

RESULT 13  
US-08-895-601-3/c  
; Sequence 3, Application US/08895601  
; Patent No. 6060262  
; GENERAL INFORMATION:  
; APPLICANT: Beer-Romero, Peggy  
; APPLICANT: Streck, Peter J.  
; APPLICANT: Glass, Susan J.  
; APPLICANT: Rolfe, Mark  
; TITLE OF INVENTION: REGULATION OF KAPPA B (1kB) DEGRADATION,  
; AND METHODS AND REAGENTS RELATED THERETO  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/895,601  
;; FILING DATE: 16-JUL-1997  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Vincent, Matthew P.  
;; REGISTRATION NUMBER: 36,709  
;; REFERENCE/DOCKET NUMBER: MIV-096.01  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-832-1000  
;; TELEFAX: 617-832-7000  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1550 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: both  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 95..1045  
US-08-895-601-3

Query Match 69.1%; Score 15.2; DB 3; Length 1550;  
Best Local Similarity 85.0%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 cgagtgctcagctagaccagc 22  
| ||||| ||||| |||||  
Db 917 CTAGTGTGAGCTGCCGAGC 898

RESULT 14  
US-08-933-750C-73/c  
; Sequence 73, Application US/08933750C  
; Patent No. 5932442  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750C  
; FILING DATE: September 23, 1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166



; TELEX:  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2028 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: UTRSNOT05  
; CLONE: 1568361  
US-08-933-750C-73

Query Match 66.4%; Score 14.6; DB 2; Length 2028;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aacgaggtgtcagctagaccag 21  
| | | | | | | | | | | | | | |  
Db 1486 ATCCAGTGTCTCAGCAACACCAG 1466

RESULT 15  
US-09-234-613-73/c  
; Sequence 73, Application US/09234613  
; Patent No. 6132973  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/234,613  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/933,750  
; FILING DATE: September 23, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0356 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2028 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: UTRSNOT05  
; CLONE: 1568361

US-09-234-613-73

Query Match 66.4%; Score 14.6; DB 3; Length 2028;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 aacgaggtgtcagctagaccag 21  
| | | | | | | | | | | | | | |  
Db 1486 ATCCAGTGTCTCAGCAACACCAG 1466

Search completed: February 25, 2002, 18:05:55  
Job time: 18603 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:20 ; Search time 8261.74 Seconds  
(without alignments)  
28.615 Million cell updates/sec

Title: US-09-698-903B-13

Perfect score: 22

Sequence: 1 aacgagtgtagtagcaccagc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	85.5	152	10	BE520593
2	18.8	85.5	174	10	BE523925
3	18.8	85.5	182	10	BE524190
4	18.8	85.5	206	10	BE525280
5	18.8	85.5	212	10	BE523695
6	18.8	85.5	223	10	BE524251
7	18.8	85.5	225	10	BE523875
8	18.8	85.5	231	10	BE522306
9	18.8	85.5	232	10	BE520764
10	18.8	85.5	232	10	BE525569
11	18.8	85.5	238	10	BE525596
12	18.8	85.5	239	10	BE523203

13	18.8	85.5	243	10	BE525564
14	18.8	85.5	244	10	BE525454
15	18.8	85.5	244	10	BE525533
16	18.8	85.5	245	10	BE525453
17	18.8	85.5	247	10	BE523445
18	18.8	85.5	249	10	BE525542
19	18.8	85.5	252	10	BE524837
20	18.8	85.5	253	10	BE525432
21	18.8	85.5	253	10	BE525500
22	18.8	85.5	254	10	BE522200
23	18.8	85.5	254	10	BE524936
24	18.8	85.5	255	10	BE525465
25	18.8	85.5	261	10	BE522627
26	18.8	85.5	272	10	BE523398
27	18.8	85.5	277	10	BE524178
28	18.8	85.5	280	10	BE520917
29	18.8	85.5	283	10	BE520827
30	18.8	85.5	287	10	BE522480
31	18.8	85.5	291	10	BE521713
32	18.8	85.5	298	10	BE523183
33	18.8	85.5	300	10	BE525377
34	18.8	85.5	303	10	BE525401
35	18.8	85.5	305	10	BE520824
36	18.8	85.5	305	10	BE521646
37	18.8	85.5	305	10	BE524379
38	18.8	85.5	306	10	BE522847
39	18.8	85.5	306	10	BE523254
40	18.8	85.5	307	10	BE523644
41	18.8	85.5	308	10	BE524918
42	18.8	85.5	310	10	BE520427
43	18.8	85.5	311	10	BE520918
44	18.8	85.5	312	10	BE520556
45	18.8	85.5	312	10	BE522759

#### ALIGNMENTS

RESULT 1

BE520593

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BE520593 152 bp mRNA EST 19-MAR-2001  
M13F7STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
clone M13F7 5', mRNA sequence.  
BE520593  
BE520593.1 GI:9778571  
EST.  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 152)  
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de  
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.  
A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
Plant Physiol. 124 (4), 1582-1594 (2000)  
20567808  
Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Michigan State University DNA Sequencing Facility Arabidopsis  
Biological Resource Center, The Ohio State University, 309 Botany &  
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
6142920603 TEL: 6142929371.  
Location/Qualifiers  
1. .152  
/organism="Arabidopsis thaliana"

FEATURES  
source

```

/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F7"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
34 t
BASE COUNT      29 a 47 c 42 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 152;
Best Local Similarity 90.9%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 2
BE523925 174 bp mRNA EST 19-MAR-2001
LOCUS M3F1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M43F11 5', mRNA sequence.
ACCESSION BE523925
VERSION BE523925.1 GI:9781903
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 174)
AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .174
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F11"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t
BASE COUNT      36 a 52 c 40 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 174;
Best Local Similarity 90.9%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

FEATURES
source
Location/Qualifiers
1. .174
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F11"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t
BASE COUNT      36 a 52 c 40 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 152;
Best Local Similarity 90.9%; Pred. No. 75;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 3
BE524190 182 bp mRNA EST 19-MAR-2001
LOCUS M47A2STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M47A2 5', mRNA sequence.
ACCESSION BE524190
VERSION BE524190.1 GI:9782168
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 182)
AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .182
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M47A2"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
42 t
BASE COUNT      33 a 58 c 49 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 182;
Best Local Similarity 90.9%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

FEATURES
source
Location/Qualifiers
1. .182
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M47A2"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
42 t
BASE COUNT      33 a 58 c 49 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 182;
Best Local Similarity 90.9%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 4
BE525280 206 bp mRNA EST 19-MAR-2001
LOCUS M61A10STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M61A10 5', mRNA sequence.
ACCESSION BE525280
VERSION BE525280.1 GI:9783258
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 206)
AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1. .206
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M61A10"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t
BASE COUNT      36 a 52 c 40 g
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 174;
Best Local Similarity 90.9%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

FEATURES
source
Location/Qualifiers
1. .174
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43F11"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
46 t
BASE COUNT      36 a 52 c 40 g
ORIGIN

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis; 1 (bases 1 to 206)  
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.  
 A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
 Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808  
 Contact: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel.: 517 355 1609  
 Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

# FEATURES

source

1..206  
 Location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="M61A10"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"  
 /note="Organ: Developing seed; Vector: pBluescript SK-;  
 Site\_1: ECORI; Site\_2: XhoII"  
 44 a 58 c 54 g 50 t

# BASE COUNT

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 206;  
 Best Local Similarity 90.9%; Pred. No. 82;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 121 AACGAGTGCACGCTGACGACG 142

# RESULT

BE523695

LOCUS

DEFINITION M40E7STM Arabidopsis developing seed Arabidopsis thaliana cDNA

ACCESSION BE523695

VERSION BE523695.1

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis; 1 (bases 1 to 212)  
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.  
 A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
 Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808  
 Contact: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel.: 517 355 1609

# REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

# FEATURES

source

1..212  
 Location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="M40E7"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"  
 /note="Organ: Developing seed; Vector: pBluescript SK-;  
 Site\_1: ECORI; Site\_2: XhoII"  
 42 a 73 c 45 g 52 t

# BASE COUNT

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 212;  
 Best Local Similarity 90.9%; Pred. No. 83;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 24 AACGAGTGCACGCTGACGACG 45

# RESULT

BE524251

LOCUS

DEFINITION M47G12STM Arabidopsis developing seed Arabidopsis thaliana cDNA

ACCESSION BE524251

VERSION BE524251.1

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis; 1 (bases 1 to 223)  
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.  
 A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil  
 Plant Physiol. 124 (4), 1582-1594 (2000)  
 20567808  
 Contact: Benning, C  
 Dept. of Biochemistry & Molecular Biology  
 Michigan State University  
 224 Biochemistry, Michigan State University, East Lansing, MI 48824  
 , USA  
 Tel.: 517 355 1609  
 Fax: 517 353 9334  
 Email: benning@msu.edu  
 Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.

JOURNAL

MEDLINE

COMMENT

# FEATURES

source

1..223  
 Location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="M47G12"  
 /clone\_lib="Arabidopsis developing seed"  
 /tissue\_type="seed"  
 /dev\_stage="5-13 days after flowering"  
 /lab\_host="E.coli"

```

/Note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: ECoRI; Site_2: XhoII"
BASE COUNT      37 a 69 c 57 g 60 t
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 223;
Best Local Similarity 90.9%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
||||| ||||| ||||| |||||
Db 99 AACGAGTCCAGCTCGACCAGC 120

RESULT 7
LOCUS BE523875 225 bp mRNA EST 19-MAR-2001
DEFINITION M43A1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE523875
VERSION BE523875.1 GI:9781853
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 225)
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1..225
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M43A1"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/Note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: ECoRI; Site_2: XhoII"
BASE COUNT      48 a 68 c 52 g 56 t
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 225;
Best Local Similarity 90.9%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
||||| ||||| ||||| |||||
Db 118 AACGAGTCCAGCTCGACCAGC 139

RESULT 8

```

```

BE522306
LOCUS BE522306 231 bp mRNA EST 19-MAR-2001
DEFINITION M25A1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE522306
VERSION BE522306.1 GI:9780284
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 231)
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers
1..231
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M25A1"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/Note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: ECoRI; Site_2: XhoII"
BASE COUNT      45 a 72 c 57 g 57 t
ORIGIN

Query Match      85.5%; Score 18.8; DB 10; Length 231;
Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgtcagctagaccagc 22
||||| ||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 9
LOCUS BE520764 232 bp mRNA EST 19-MAR-2001
DEFINITION M14H9STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE520764
VERSION BE520764.1 GI:9778742
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 232)
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
A new set of Arabidopsis expressed sequence tags from developing

```

```

seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)

JOURNAL MEDLINE COMMENT
20567808
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.

FEATURES
source
1. .232
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600014455R1"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
44 a 73 c 57 g 58 t

Query Match 85.5%; Score 18.8; DB 10; Length 232;
Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||||| ||||| |||||
Db 57 AACGAGTCCAGCTCGACCAGC 78

RESULT 10
BE525569
LOCUS M62M13STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone 600014455R1 5', mRNA sequence.
ACCESSION BE525569
VERSION BE525569.1 GI:9783470
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 232)
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Javorski,J.G., Ohlrogge,J. and Benning,C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES
source
1. .238
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600014490R1"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
45 a 70 c 63 g 58 t 2 others

Query Match 85.5%; Score 18.8; DB 10; Length 238;

```

```

1. .232
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600014455R1"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
41 a 72 c 56 g 63 t

BASE COUNT
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 232;
Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgcagctagaccagc 22
||||||| ||||| |||||
Db 90 AACGAGTCCAGCTCGACCAGC 111

RESULT 11
BE525596
LOCUS M62C23STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone 600014490R1 5', mRNA sequence.
ACCESSION BE525596
VERSION BE525596.1 GI:9783497
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 238)
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Javorski,J.G., Ohlrogge,J. and Benning,C.
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.

FEATURES
source
1. .238
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600014490R1"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"
45 a 70 c 63 g 58 t 2 others

BASE COUNT
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 238;

```

Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 aacgagtgcagctagaccagc 22  
||||| ||||| ||||| |||||  
Db 24 AACGAGTCCAGCTCGACCAGC 45

RESULT 12  
BE523203  
LOCUS M35C7STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
DEFINITION clone M35C7 5', mRNA sequence.  
ACCESSION BE523203  
VERSION BE523203.1 GI:9781277  
KEYWORDS EST.

SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 239)  
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)  
MEDLINE 20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142929371.

FEATURES  
source  
1..239  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="M35C7"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
BASE COUNT 49 a 74 c 57 g 59 t  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 239;  
Best Local Similarity 90.9%; Pred. No. 85;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 aacgagtgcagctagaccagc 22  
||||| ||||| ||||| |||||  
Db 113 AACGAGTCCAGCTCGACCAGC 134

RESULT 13  
BE525564  
LOCUS M6213STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
DEFINITION clone 600014450R1 5', mRNA sequence.  
ACCESSION BE525564  
VERSION BE525564.1 GI:9783465  
KEYWORDS EST.

SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 243)  
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)  
MEDLINE 20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:  
6142929371.

SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES  
source

1..243  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="600014450R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: EcoRI; Site\_2: XhoII"  
BASE COUNT 49 a 69 c 65 g 59 t  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 243;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 aacgagtgcagctagaccagc 22  
||||| ||||| ||||| |||||  
Db 21 AACGAGTCCAGCTCGACCAGC 42

RESULT 14

BE525454

LOCUS M62603STM Arabidopsis developing seed Arabidopsis thaliana cDNA

DEFINITION clone 600014412R1 5', mRNA sequence.

ACCESSION BE525454

VERSION BE525454.1 GI:9783432

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 244)  
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de  
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.  
TITLE A new set of Arabidopsis expressed sequence tags from developing  
seeds. The metabolic pathway from carbohydrates to seed oil  
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)  
MEDLINE 20567808  
COMMENT Contact: Benning, C  
Dept. of Biochemistry & Molecular Biology  
Michigan State University  
224 Biochemistry, Michigan State University, East Lansing, MI 48824  
, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142929371.



, USA  
Tel: 517 355 1609  
Fax: 517 353 9334  
Email: benning@msu.edu  
Clones were originally prepared at Michigan State University.  
Arabidopsis Biological Resource Center, The Ohio State University,  
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210  
USA, FAX: 6142920603 TEL: 6142929371.

## FEATURES

source

1. .244  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone\_lib="600014412R1"  
/clone\_lib="Arabidopsis developing seed"  
/tissue\_type="seed"  
/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: ECORI; Site\_2: XhoII"

BASE COUNT 46 a 74 c 66 g 57 t 1 others  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 244;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| |||||  
Db 44 AACGAGTCCAGCTCGACCAGC 65

## RESULT 15

BE525533

LOCUS

DEFINITION BE525533 244 bp mRNA EST 19-MAR-2001  
M62K03STM Arabidopsis developing seed Arabidopsis thaliana cDNA  
clone 600014414R1 5', mRNA sequence.

ACCESSION

BE525533

VERSION

BE525533.1

GI:9783434

KEYWORDS

EST

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 244)

White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de

Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.

A new set of Arabidopsis expressed sequence tags from developing

seeds. The metabolic pathway from carbohydrates to seed oil

Plant Physiol. 124 (4), 1582-1594 (2000)

20567808

Contact: Benning, C

Dept. of Biochemistry &amp; Molecular Biology

Michigan State University

224 Biochemistry, Michigan State University, East Lansing, MI 48824

, USA

Tel: 517 355 1609

Fax: 517 353 9334

Email: benning@msu.edu

Clones were originally prepared at Michigan State University.

Arabidopsis Biological Resource Center, The Ohio State University,

309 Botany &amp; Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210

USA, FAX: 6142920603 TEL: 6142929371.

Location/Qualifiers

1. .244

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone\_lib="600014414R1"

/clone\_lib="Arabidopsis developing seed"

/tissue\_type="seed"

/dev\_stage="5-13 days after flowering"  
/lab\_host="E.coli"  
/note="Organ: Developing seed; Vector: pBluescript SK-;  
Site\_1: ECORI; Site\_2: XhoII"  
BASE COUNT 43 a 74 c 60 g 67 t  
ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 244;  
Best Local Similarity 90.9%; Pred. No. 86;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aacgagtgtcagctagaccagc 22  
||||||| ||||| |||||  
Db 91 AACGAGTCCAGCTCGACCAGC 112

Search completed: February 25, 2002, 17:21:21  
Job time: 16174 sec

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:03:03 ; Search time 2331.3 Seconds  
(without alignments)  
155.680 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22  
Sequence: 1 cgcagtcctgtgaacatcgacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.ov.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo\_hum.\*
- 31: em.htgo\_inv.\*
- 32: em.htgo\_rod.\*
- 33: em.htg\_hum.\*
- 34: em.htg\_inv.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	6	AX127761
2	22	100.0	22	6	AX172482
3	22	100.0	3113	6	I24540 Sequence 20
4	22	100.0	3113	6	I83673 Sequence 2
5	22	100.0	3113	8	X14555 Brassica na
6	22	100.0	3198	8	X59294 B.napus Bnc
7	17.4	79.1	143830	33	AC021599 Homo sapi
8	17.4	79.1	151210	2	AC092613 Homo sapi
9	17.4	79.1	154016	9	AC005047 Homo sapi
10	17.2	78.2	287	4	AF142615 Equus cab
11	17.2	78.2	587	6	I56095 Sequence 3
12	17.2	78.2	646	1	AF190914 Escherich
13	17.2	78.2	1174	6	I56097 Sequence 5
14	17.2	78.2	1188	6	I56101 Sequence 9
15	17.2	78.2	1196	6	I56099 Sequence 7
16	17.2	78.2	1264	1	X05952 Shigella so
17	17.2	78.2	1433	1	U97492 Shigella fl
18	17.2	78.2	3294	1	M24352 S.dysenter
19	17.2	78.2	3715	1	AF081284 Escherich
20	17.2	78.2	3719	1	U81136 Shigella fl
21	17.2	78.2	5430	1	U82621 Shigella fl
22	17.2	78.2	6014	1	AJ271153 Shigella
23	17.2	78.2	7150	1	AF177050 Shigella
24	17.2	78.2	10029	1	AE007213 Sinorhizo
25	17.2	78.2	10040	1	AE005271 Escherich
26	17.2	78.2	10605	1	AF139596 Shigella
27	17.2	78.2	11058	1	AE005307 Escherich
28	17.2	78.2	15708	3	AF051097 Balanoglo
29	17.2	78.2	20962	1	AF335540 Shigella
30	17.2	78.2	20962	1	AF335540 Shigella
31	17.2	78.2	23771	1	AF141323 Shigella
32	17.2	78.2	32094	1	AF153317 Shigella
33	17.2	78.2	37010	1	EC0278144
34	17.2	78.2	51334	1	AF200692 Shigella
35	17.2	78.2	77283	2	AC020099 Drosophil
36	17.2	78.2	92077	1	AF074613 Escherich
37	17.2	78.2	92077	6	AX191727 Sequence
38	17.2	78.2	92721	1	AB011549 Escherich
39	17.2	78.2	92721	6	AX191725 Sequence
40	17.2	78.2	165190	3	AC093096 Drosophil
41	17.2	78.2	213494	1	SFPWR100 Shigella
42	17.2	78.2	213494	1	SFPWR100 Shigella
43	17.2	78.2	217280	2	AC019337 Homo sapi
44	17.2	78.2	221851	1	AF348706 Shigella
45	17.2	78.2	221851	1	AF348706 Shigella

ALIGNMENTS

RESULT 1					
AX127761	AX127761	Sequence 14	22 bp	DNA	PAT 15-MAY-2001
LOCUS	Sequence 14	from Patent	WO0131042.		
DEFINITION	AX127761				
ACCESSION	AX127761.1	GI:14134408			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
synthetic construct.					
artificial construct					
1 (bases 1 to 22)					
REFERENCE					
AUTHORS	Weston,B. and de Beuckeleer,M.				
TITLE	Male-sterile brassica plants and methods for producing same				
JOURNAL	Patent: WO 0131042-A 14 03-MAY-2001;				
AVANTIS CropScience N.V. (BE)					
Location/Qualifiers					
1..22					
/organism="synthetic construct"					
/db_xref="taxon:32630"					
/note="primer CV28"					

BASE COUNT 5 a 7 c 5 g 5 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1 CGCAGTTCTGTGACATCGACC 22

## RESULT 2

AX172482  
LOCUS AX172482 22 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 43 from Patent WO0141558.  
ACCESSION AX172482  
VERSION AX172482.1 GI:14597594  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM  
REFERENCE 1 (bases 1 to 22)  
AUTHORS de Both, G. and de Beuckeleer, M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 43 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES  
source Location/Qualifiers  
1..22  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer BNA06"

BASE COUNT 5 a 7 c 5 g 5 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1 CGCAGTTCTGTGACATCGACC 22

## RESULT 3

I24540/c  
LOCUS I24540 3113 bp DNA PAT 07-OCT-1996  
DEFINITION Sequence 20 from patent US 5543576.  
ACCESSION I24540  
VERSION I24540.1 GI:1604410  
KEYWORDS  
SOURCE Unknown.  
ORGANISM  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS van Ooijen, A.J.J., Rietveld, K., Hoekema, A., Pen, J., Sijmons, P.C.,  
Verwoerd, R.C., and Quax, W.J.  
TITLE Production of enzymes in seeds and their use  
JOURNAL Patent: US 5543576-A 20 06-AUG-1996;  
MEDLINE  
source Location/Qualifiers  
1..3113  
/organism="unknown"

BASE COUNT 961 a 685 c 586 g 881 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22

Db 1183 CGCAGTTCTGTGACATCGACC 1162  
|||||

## RESULT 4

I83673/c  
LOCUS I83673 3113 bp DNA PAT 10-AUG-1998  
DEFINITION Sequence 2 from patent US 5714474.  
ACCESSION I83673  
VERSION I83673.1 GI:3407203  
KEYWORDS  
SOURCE Unknown.  
ORGANISM  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS Van Ooijen, A.J.J., Rietveld, K., Hoekema, A., Pen, J.,  
Sijmons, P.C., Christian, Verwoerd, T., Cornelis and Quax, W. Johannes.  
TITLE Production of enzymes in seeds and their use  
JOURNAL Patent: US 5714474-A 2 03-FEB-1998;  
FEATURES  
source Location/Qualifiers  
1..3113  
/organism="unknown"

BASE COUNT 961 a 684 c 586 g 882 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
|||||  
Db 1183 CGCAGTTCTGTGACATCGACC 1162

## RESULT 5

BNCRUA/c  
LOCUS BNCRUA 3113 bp DNA PLN 10-FEB-1999  
DEFINITION Brassica napus crUA gene for cruciferin.  
ACCESSION X14555  
VERSION X14555.1 GI:17810  
KEYWORDS crUA gene; cruciferin; seed storage protein.  
SOURCE rape.  
ORGANISM Brassica napus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 3113)  
AUTHORS Ryan, A.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAR-1989) Ryan A.J., Department of Biological  
Sciences, University of Durham, South Road, Durham, DH1 3LE, United  
Kingdom  
REFERENCE 2 (bases 1 to 3113)  
AUTHORS Ryan, A.J., Royal, C.L., Hutchinson, J., and Shaw, C.H.  
TITLE Genomic sequence of a 12S seed storage protein from oilseed rape  
JOURNAL (Brassica napus c.v. Jet Neuf)  
MEDLINE Nucleic Acids Res. 17 (9), 3584 (1989)  
source Location/Qualifiers  
1..3113  
/organism="Brassica napus"  
/strain="Jet Neuf"  
/db\_xref="taxon:3708"  
/tissue.type="leaf"  
644..650  
/note="pot. TATA-box"  
680..3010  
/note="primary transcript"  
Join(709..991,1220..1581,2049..2870)  
/codon\_start=1  
/product="cruciferin"  
/protein\_id="CAA32692.1"

promoter

precursor\_RNA

CDS

```
/db_xref="GI:17811"
/db_xref="SWISS-PROT:P11090"
/translation="MARLSLILSFLLALFLHGSTAQQFPNECQDQLNALPESHVL
KAEGRIEVDHAPQLRCGVSFVRYLIESKGLYLSPFFSTALSPVAKGEGLMGRV
VLCATETQDSVFPQSGSPGEGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
FRDMHQVHEIRTDGTIATHPGVAOMFYNDGNQPLVIVSLDLASHQNLDRNRPFF
LAGNPGQGVWIEGREQQPKNIILNGFTPEVLAKAFKIDVRTAQQLNQDNRGNII
VQGFPSVIRPRLRSQRPOETEVNGLEETICSARCTNLDLDDPSNADVYKPOLGYISTL
YDLPILRLRLSALRGSIRQNAWLPQWNAWLVVTDGEAHVQVVDNGDRVFDG
VQSGQQLLSIPQGSFVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVIS
GYQISLEEARVKENTETITLTHSSGFASGGPKRADA"
992..1219
/note="intron I"
1582..2048
/note="intron II"
2915..2920
/note="polyA signal"
2989..2994
/note="polyA signal"
BASE COUNT 961 a 685 c 586 g 881 t
ORIGIN
intron
intron
misc_feature
misc_feature
Query Match 100.0%; Score 22; DB 8; Length 3113;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCTGTGACATCGACC 1162

RESULT 6
BNC1G/C BNC1G BNC1G 3198 bp DNA PLN 04-APR-1995
LOCUS B.napus BNC1 gene for cruciferin storage protein.
DEFINITION X59294
ACCESSION X59294
VERSION X59294.1 GI:17790
KEYWORDS cruciferin; cruciferin storage protein.
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 3198)
Breen,J.P. and Crouch,M.L.
Molecular analysis of a cruciferin storage protein gene family of
Brassica napus
Plant Mol. Biol. 19 (6), 1049-1055 (1992)
92379259
2 (bases 1 to 3198)
Breen,J.P.
Direct Submission
Submitted (24-APR-1991) Breen J.P., Jordan Hall, Indiana
University, Bloomington, IN 47401, USA
See also M16860 & X59295 (for Bnc2 gene).
FEATURES
source
1..3198
/organism="Brassica napus"
/strain="C.V. Tower"
/db_xref="taxon:3708"
/cell_line="ED8767"
/clone_lib="Ch4a"
/clone="lambda Bnc1"
join(<709..991,1220..1584,2069..2488,2653..>3057)
/gene="Bnc1"
709..3057
/gene="Bnc1"
join(709..991,1220..1584,2069..2488,2653..3057)
/gene="Bnc1"
/codon_start=1
/product="cruciferin storage protein"
/protein_id="CAA41984.1"
```

```
/db_xref="GI:762919"
/db_xref="SWISS-PROT:P33523"
/translation="MARLSLILSFLLALFLHGSTAQQFPNECQDQLNALPESHVL
KAEGRIEVDHAPQLRCGVSFVRYLIESKGLYLSPFFSTALSPVAKGEGLMGRV
VPGCAETQDSVFPQSGSPGEGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQ
FRDMHQVHEIRTDGTIATHPGVAOMFYNDGNQPLVIVSLDLASHQNLDRNRPFF
YLAGNPGQGVWIEGREQQPKNIILNGFTPEVLAKAFKIDVRTAQQLNQDNRGNII
RVQGFPSVIRPRLRSQRPOETEVNGLEETICSARCTNLDLDDPSNADVYKPOLGYISTL
NSYDLPILRLRLSALRGSIRQNAWLPQWNAWLVVTDGEAHVQVVDNGDRVFDG
VQSGQQLLSIPQGSFVVKRATSEQFRWIEFKTNANAQINTLAGRTSVLRGLPLEVI
SNGQISLEEARVKENTETITLTHSSGFASGGPKRADA"
<709..991
/gene="Bnc1"
/number=1
992..1219
/note="intron"
/gene="Bnc1"
/number=1
1220..1584
/gene="Bnc1"
/number=2
1585..2068
/note="intron"
/gene="Bnc1"
/number=2
2069..2488
/gene="Bnc1"
/number=3
2489..2652
/note="intron"
/gene="Bnc1"
/number=3
2654..>3057
/note="Bnc1"
/number=4
polyA_signal 3102..3107
polyA_signal 3176..3181
BASE COUNT 992 a 690 c 600 g 916 t
ORIGIN
Query Match 100.0%; Score 22; DB 8; Length 3198;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 cgcagttctgtgaacatcgacc 22
|||||
Db 1183 CGCAGTTCTGTGACATCGACC 1162

RESULT 7
AC021599
ID AC021599 standard; DNA; HTG; 143830 BP.
XX AC021599;
SV AC021599.3
XX
DT 17-JAN-2000 (Rel. 62, Created)
DT 21-SEP-2000 (Rel. 65, Last updated, Version 3)
XX
DE Homo sapiens clone RP11-189E18, WORKING DRAFT SEQUENCE, 29 unordered
DE pieces.
DE
DE HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP 1-143830
RA Birren B., Linton L., Nusbaum C., Lander E.;
RT "Homo sapiens, clone RP11-189E18";
RL Unpublished.
XX
```

```

[2]
RN 1-143830
RP Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
RA Anderson S., Baldwin J., Barna N., Beckerly R., Bada F., Boguslavskiy L.,
RA Boukhgalter B., Brown A., Burkett G., Castle A., Choepel Y., Colangelo M.,
RA Collins S., Collymore A., Cooke P., DeArellano K., Dewar K., Domino M.,
RA Doyle M., Fenestor J., Ferreira P., FitzHugh W., Forrest C., Gage D.,
RA Galagan J., Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,
RA Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,
RA Landers T., Lehoczyk J., Levine R., Lieu C., Liu G., Locke K.,
RA Macdonald P., Marquis N., McEwan P., McGuirk A., McKernan K., McPheeters R.,
RA Meldrim J., Meneus L., Morrow J., Naylor J., Norman C.H., O'Connor T.,
RA O'Donnell P., Oliver T.M., Peterson K., Pierre N., Pisanl C., Pollara V.,
RA Raymond C., Riley R., Rothman D., Roy A., Santos R., Severy P., Spencer B.,
RA Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J., Testaye S.,
RA Theodore J., Tirrell A., Vassiliev H., Viel R., Vo A., Wu X., Wyman D.,
RA Ye W.J., Zimmer A., Zody M.;
RT
RL
RL Submitted (16-JAN-2000) to the EMBL/GenBank/DBJ databases.
RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL Cambridge, MA 02141, USA
XX
CC On Sep 20, 2000 this sequence version replaced gi:9152655.
CC All repeats were identified using RepeatMasker:
CC Smit, A.F.A. & Green, P. (1996-1997)
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC ----- Genome Center
CC Center: Whitehead Institute/ MIT Center for Genome Research
CC Center code: WIRK
CC Web site: http://www-seq.wi.mit.edu
CC Contact: sequence\_submissions@genome.wi.mit.edu
CC ----- Project Information
CC Center project name: L5793
CC Center clone name: 189_E_18
CC ----- Summary Statistics
CC Sequencing vector: M13; M77815; 100% of reads
CC Chemistry: Dye-terminator Big Dye; 100% of reads
CC Assembly program: Phrap; version 0.960731
CC Consensus quality: 130936 bases at least Q40
CC Consensus quality: 137518 bases at least Q30
CC Consensus quality: 139807 bases at least Q20
CC Insert size: 150000; agarose-fp
CC Quality coverage: 3.7 in Q20 bases; agarose-fp
CC Quality coverage: 3.9 in Q20 bases; sum-of-contigs
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC * consists of 29 contigs. The true order of the pieces
CC * is not known and their order in this sequence record is
CC * arbitrary. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence
CC * as soon as it is available and the accession number will
CC * be preserved.
CC
CC 1 830: contig of 830 bp in length
CC * 831 930: gap of 100 bp
CC * 931 2470: contig of 1540 bp in length
CC * 2471 2570: gap of 100 bp
CC * 2571 4343: contig of 1773 bp in length
CC * 4344 4443: gap of 100 bp
CC * 4444 5565: contig of 1122 bp in length
CC * 5566 5665: gap of 100 bp
CC * 5666 7543: contig of 1878 bp in length
CC * 7544 7643: gap of 100 bp
CC * 7644 9898: contig of 2255 bp in length
CC * 9899 9998: gap of 100 bp
CC * 9999 11795: contig of 1797 bp in length
CC * 11796 11895: gap of 100 bp
CC * 11896 14229: contig of 2334 bp in length
CC * 14230 14329: gap of 100 bp
CC * 14330 16423: contig of 2094 bp in length
CC * 16424 16523: gap of 100 bp
CC * 16524 19064: contig of 2541 bp in length

```

```

CC * 19065 19164: gap of 100 bp
CC * 19165 21284: contig of 2120 bp in length
CC * 21285 21384: gap of 100 bp
CC * 21385 23749: contig of 2365 bp in length
CC * 23750 23849: gap of 100 bp
CC * 23850 27606: contig of 3757 bp in length
CC * 27607 27706: gap of 100 bp
CC * 27707 31136: contig of 3430 bp in length
CC * 31137 31236: gap of 100 bp
CC * 31237 35355: contig of 4119 bp in length
CC * 35356 35455: gap of 100 bp
CC * 35456 39501: contig of 4046 bp in length
CC * 39502 39601: gap of 100 bp
CC * 39602 44277: contig of 4676 bp in length
CC * 44278 44377: gap of 100 bp
CC * 44378 47808: contig of 3431 bp in length
CC * 47809 47908: gap of 100 bp
CC * 47909 52253: contig of 4345 bp in length
CC * 52254 52353: gap of 100 bp
CC * 52354 57551: contig of 5198 bp in length
CC * 57552 57651: gap of 100 bp
CC * 57652 64108: contig of 6457 bp in length
CC * 64109 64208: gap of 100 bp
CC * 64209 71334: contig of 7126 bp in length
CC * 71335 71434: gap of 100 bp
CC * 71435 79411: contig of 7977 bp in length
CC * 79412 79511: gap of 100 bp
CC * 79512 86611: contig of 7100 bp in length
CC * 86612 86711: gap of 100 bp
CC * 86712 94930: contig of 8219 bp in length
CC * 94931 95030: gap of 100 bp
CC * 95031 101723: contig of 6693 bp in length
CC * 101724 101823: gap of 100 bp
CC * 101824 116518: contig of 14695 bp in length
CC * 116519 116618: gap of 100 bp
CC * 116619 129081: contig of 12463 bp in length
CC * 129082 129181: gap of 100 bp
CC * 129182 143830: contig of 14649 bp in length.
XX
FH
FH Key Location/Qualifiers
FH Source 1. .143830
FH /db_xref="taxon:9606"
FH /organism="Homo sapiens"
FH /clone="Rp11-189E18"
FH /clone_lib="RPC1-11 Human Male BAC"
FH misc_feature 1. .830
FH /note="assembly_fragment clone_end:SP6 vector_side:left"
FH misc_feature 931. .2470
FH /note="assembly_fragment"
FH misc_feature 2571. 4343
FH /note="assembly_fragment"
FH misc_feature 4444. .5565
FH /note="assembly_fragment"
FH misc_feature 5566. .7543
FH /note="assembly_fragment"
FH misc_feature 7644. 9898
FH /note="assembly_fragment"
FH misc_feature 9999. .11795
FH /note="assembly_fragment"
FH misc_feature 11896. .14229
FH /note="assembly_fragment"
FH misc_feature 14330. .16423
FH /note="assembly_fragment"
FH misc_feature 16524. .19064
FH /note="assembly_fragment"
FH misc_feature 19165. .21284
FH /note="assembly_fragment"
FH misc_feature 21385. .23749
FH /note="assembly_fragment"
FH misc_feature 23850. .27606
FH /note="assembly_fragment"
FH misc_feature 27707. .31136

```

FT misc\_feature /note="assembly\_fragment"  
31237. .35355  
FT /note="assembly\_fragment"  
35456. .39501  
FT /note="assembly\_fragment"  
39602. .44277  
FT /note="assembly\_fragment"  
44378. .47808  
FT /note="assembly\_fragment"  
47909. .52253  
FT /note="assembly\_fragment"  
52354. .57551  
FT /note="assembly\_fragment"  
57652. .64108  
FT /note="assembly\_fragment"  
64209. .71334  
FT /note="assembly\_fragment"  
71435. .79411  
FT /note="assembly\_fragment"  
79512. .86611  
FT /note="assembly\_fragment"  
86712. .94930  
FT /note="assembly\_fragment"  
95031. .101723  
FT /note="assembly\_fragment"  
101824. .116518  
FT /note="assembly\_fragment"  
116619. .129081  
FT /note="assembly\_fragment"  
129182. .143830

Query Match 79.1%; Score 17.4; DB 33; Length 143830;  
Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
|||||  
Db 83635 GCAGTTCTGTGAACATAGA 83653

RESULT 8  
AC092613/c  
LOCUS AC092613 151210 bp DNA HTG 19-JUL-2001  
DEFINITION Homo sapiens chromosome 7 clone RP11-189E18, WORKING DRAFT  
SEQUENCE, 1 unordered pieces.  
ACCESSION AC092613 AC021599  
VERSION AC092613.1 GI:14916198  
KEYWORDS HTG; HTGS\_PHASE1; HTGS-DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 151210)  
AUTHORS Waterston,R.H.  
JOURNAL The sequence of Homo sapiens clone  
Unpublished  
REFERENCE 2 (bases 1 to 151210)  
AUTHORS Waterston,R.H.  
JOURNAL Direct Submission  
TITLE Waterston,R.H.  
Submitted (19-JUL-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Jul 19, 2001 this sequence version replaced gi:10198499.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0189E18  
Drafting center: WIBR  
----- Summary Statistics -----

Sequencing vector: M13; 26%  
Chemistry: Dye-primer ET; 37%  
Chemistry: Dye-terminator Big Dye; 63% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 148642 bases at least Q40  
Consensus quality: 149695 bases at least Q30  
Consensus quality: 151204 bases at least Q20  
Insert size: 143000; agarose-fp  
Quality coverage: 8.04 in Q20 bases; agarose-fp  
Quality coverage: 7.60 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 151210: contig of 151210 bp in length.  
\* Location/Qualifiers  
\* 1..151210  
\* /organism="Homo sapiens"  
\* /db\_xref="taxon:9606"  
\* /chromosome="7"  
\* /clone="RP11-189E18"  
\* 1..151210  
\* /note="assembly\_name:Contig5  
\* clone\_end:SP6  
\* vector\_side:right"

FEATURES  
source

misc\_feature  
1..151210  
/note="assembly\_name:Contig5  
clone\_end:SP6  
vector\_side:right"

BASE COUNT 45173 a 28928 c 30068 g 47041 t  
ORIGIN

Query Match 79.1%; Score 17.4; DB 2; Length 151210;  
Best Local Similarity 94.7%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcagttctgtgaacatcga 20  
|||||  
Db 17340 GCAGTTCTGTGAACATAGA 17322

RESULT 9

AC005047  
LOCUS AC005047 154016 bp DNA PRI 07-OCT-2000  
DEFINITION Homo sapiens BAC clone CTB-14E15 from 7q22-q32, complete sequence.  
ACCESSION AC005047  
VERSION AC005047.3 GI:10716665  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 154016)  
AUTHORS Sulston,J.E. and Waterston,R.  
JOURNAL Toward a complete human genome sequence  
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
REFERENCE 2 (bases 1 to 154016)  
AUTHORS Courtney,L., Gillam,B., Stoneking,T., Elliott,G. and Langston,Y.  
JOURNAL The sequence of Homo sapiens BAC clone CTB-14E15  
Unpublished  
REFERENCE 3 (bases 1 to 154016)  
AUTHORS Waterston,R.H.  
JOURNAL Direct Submission  
TITLE Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 154016)  
AUTHORS Waterston,R.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (07-OCT-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 7, 2000 this sequence version replaced gi:7631121.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_RG014E15  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send  
<mailto:egreen@ngri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-14E15 is from the first release of the human BAC library  
CTB-978SK-B. The library contains cloned DNA from the male  
fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl.  
Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8  
(1996). This clone is available from Research Genetics, Inc.  
(<http://www.resgen.com>).  
VECTOR: pBel0BAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is GS1-34D21, 200 base pair  
overlap. Actual start of this clone is at base position 1 of  
CTB-14E15; actual end is at base position 7330 of GS1-34D21.

Due to low quality, the fidelity of the sequence from 130932 to  
130982 can not be guaranteed.

FEATURES  
source

Location/Qualifiers  
1. .154016  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q22-q32"  
/clone="CTB-14E15"  
/clone\_lib="CTB-978SK-B"  
589. .874  
/rpt\_family="L1"  
903. .2491  
/rpt\_family="ERV1"  
2495. .2700  
/rpt\_family="L1"  
2720. .2838  
/rpt\_family="Alu"  
2839. .3289  
/rpt\_family="L1"  
3411. .3476  
/rpt\_family="MaLR"

repeat\_region 3477. .3783  
/rpt\_family="Alu"  
repeat\_region 3784. .4135  
/rpt\_family="MaLR"  
repeat\_region 4849. .5153  
/rpt\_family="L1"  
repeat\_region 5740. .6039  
/rpt\_family="Alu"  
repeat\_region 6077. .6151  
/rpt\_family="MER99"  
repeat\_region 6319. .6616  
/rpt\_family="CRI"  
repeat\_region 8383. .8761  
/rpt\_family="Alu"  
repeat\_region 8771. .8906  
/rpt\_family="Alu"  
repeat\_region 9238. .9514  
/rpt\_family="Alu"  
repeat\_region 10478. .10731  
/rpt\_family="ERV1"  
STS 11164. .11523  
/db\_xref="GI:6120582"  
repeat\_region 12056. .12181  
/rpt\_family="MaLR"  
repeat\_region 12356. .12482  
/rpt\_family="MIR"  
repeat\_region 12500. .12625  
/rpt\_family="CRI"  
repeat\_region 12717. .12750  
/rpt\_family="CRI"  
repeat\_region 12817. .12901  
/rpt\_family="CRI"  
repeat\_region 13158. .13555  
/rpt\_family="ERV1"  
repeat\_region 14352. .14508  
/rpt\_family="L1"  
repeat\_region 14509. .14900  
/rpt\_family="MaLR"  
repeat\_region 14901. .15114  
/rpt\_family="L1"  
repeat\_region 15152. .15625  
/rpt\_family="ERV1"  
repeat\_region 15626. .16096  
/rpt\_family="L1"  
repeat\_region 16147. .16636  
/rpt\_family="ERV1"  
repeat\_region 16640. .17690  
/rpt\_family="ERVK"  
repeat\_region 17710. .17806  
/rpt\_family="ERV1"  
repeat\_region 17850. .17908  
/rpt\_family="L2"  
repeat\_region 18099. .18221  
/rpt\_family="MER103"  
STS 18305. .18622  
/db\_xref="GI:485206"  
repeat\_region 18833. .19179  
/rpt\_family="MER2\_type"  
repeat\_region 19215. .19509  
/rpt\_family="Alu"  
repeat\_region 19983. .20033  
/rpt\_family="L2"  
repeat\_region 20109. .20370  
/rpt\_family="L2"  
repeat\_region 20708. .20760  
/rpt\_family="MaLR"  
repeat\_region 20933. .21089  
/rpt\_family="MaLR"  
repeat\_region 21094. .21147  
/rpt\_family="ERV1"  
repeat\_region 21148. .21458  
/rpt\_family="Alu"  
repeat\_region 21459. .21613



```

source
1. .287
/organism="Equus caballus"
/db_xref="taxon:9796"
/chromosome="2"
59. .70
repeat_region
/note="microsatellite COR078"
/rpt_type=tandem
/rpt_unit=gt
56 a 102 g 92 t
BASE COUNT
ORIGIN
```

```

source      l.040
            /organism="Escherichia coli"
            /plasmid="Ent"
            /strain="K2"
            /db_xref="taxon:562"
            <1.>646
            /gene="eltB"
gene

```

BASE COUNT 182 a 134 c 149 g 175 t 6 others  
ORIGIN

/note="3' flanking sequence"

Query Match 78.2%; Score 17.2; DB 1; Length 646;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 586 CGCAGTACTGTGAACCTCGATC 607

RESULT 13  
LOCUS I56097 1174 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 5 from patent US 5648481.  
ACCESSION I56097  
VERSION I56097.1 GI:2476891

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1174)  
AUTHORS Parodos,K. and McCarty,J.  
TITLE Nucleic acid probes for the detection of shigella  
JOURNAL Patent: US 5648481-A 5 15-JUL-1997;  
FEATURES Location/Qualifiers  
source 1..1174

BASE COUNT 252 a 296 c 276 g 330 t 20 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1174;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 322 CGCAGTACTGTGAACCTCGATC 343

RESULT 14  
LOCUS I56101 1188 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 9 from patent US 5648481.  
ACCESSION I56101

VERSION I56101.1 GI:2476895  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1188)  
AUTHORS Parodos,K. and McCarty,J.  
TITLE Nucleic acid probes for the detection of shigella  
JOURNAL Patent: US 5648481-A 9 15-JUL-1997;  
FEATURES Location/Qualifiers  
source 1..1188

BASE COUNT 258 a 311 c 273 g 330 t 16 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1188;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 195 CGCAGTACTGTGAACCTCGATC 216

RESULT 15  
LOCUS I56099 1196 bp DNA PAT 07-OCT-1997  
DEFINITION Sequence 7 from patent US 5648481.  
ACCESSION I56099  
VERSION I56099.1 GI:2476893

KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1196)  
AUTHORS Parodos,K. and McCarty,J.  
TITLE Nucleic acid probes for the detection of shigella  
JOURNAL Patent: US 5648481-A 7 15-JUL-1997;  
FEATURES Location/Qualifiers  
source 1..1196

BASE COUNT 259 a 298 c 272 g 353 t 14 others  
ORIGIN

Query Match 78.2%; Score 17.2; DB 6; Length 1196;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
||||| ||||||| |||||

Db 195 CGCAGTACTGTGAACCTCGATC 216

Search completed: February 25, 2002, 18:03:11  
Job time: 18599 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:42 ; Search time 716.55 Seconds  
(without alignments)  
26.322 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22

Sequence: 1 cgcagttctgtgaacatcgacc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	22	100.0	22	AAH25458	PCR primer for end
2	22	100.0	22	AAH25458	PCR primer B04, to
3	22	100.0	3113	AAQ07003	Cruciferin A gene.
4	17.2	78.2	587	AAQ37508	Shigella specific
5	17.2	78.2	1174	AAQ37510	E. coli repeat 1.
6	17.2	78.2	1188	AAQ37514	S. flexneri repeat
7	17.2	78.2	1196	AAQ37512	E. coli repeat 2
8	16.8	76.4	2019	AAV38672	Mus musculus SCS7
9	16.8	76.4	7316	AAH21115	C. glutamicum LP-6
10	16.4	74.5	4961	AAH77036	Human ORFX ORF2591
11	15.8	71.8	226	AAC12754	Human secreted pro

c	12	15.8	71.8	319608	21	AAH51601	Human chromosome 1
c	13	15.8	71.8	319608	22	AAH51601	Human chromosome 1
	14	15.6	70.9	545	20	AAH51601	Tobacco leaf poly
	15	15.6	70.9	769	18	AAH51601	Streptococcus pneu
	16	15.6	70.9	769	19	AAH51601	DNA encoding a S.
	17	15.6	70.9	2108	22	AAH51601	Neisseria meningit
	18	15.6	70.9	2109	22	AAH51601	Neisseria meningit
	19	15.6	70.9	2111	22	AAH51601	Neisseria meningit
	20	15.6	70.9	2112	21	AAH51601	DNA encoding a Nei
	21	15.6	70.9	2112	21	AAH51601	Neisseria meningit
	22	15.6	70.9	2112	22	AAH51601	Neisseria meningit
	23	15.6	70.9	2112	22	AAH51601	Neisseria meningit
	24	15.6	70.9	2112	22	AAH51601	Neisseria meningit
	25	15.6	70.9	2112	22	AAH51601	Neisseria meningit
	26	15.6	70.9	2112	22	AAH51601	Neisseria meningit
	27	15.6	70.9	2112	22	AAH51601	Neisseria meningit
c	28	15.6	70.9	2582	7	AAH51601	Vector sequence de
c	29	15.6	70.9	2585	8	AAH51601	DNA - a sequence of
c	30	15.6	70.9	8367	21	AAH51601	N. meningitidis pa
	31	15.6	70.9	21185	21	AAH51601	Streptomyces globi
	32	15.6	70.9	63164	21	AAH51601	Streptomyces globi
	33	15.6	70.9	349980	21	AAH51601	Neisseria meningit
	34	15.6	70.9	349980	21	AAH51601	Neisseria meningit
	35	15.6	70.9	1437688	21	AAH51601	N. meningitidis B
c	36	15.4	70.0	140	21	AAH51601	Human secreted pro
c	37	15.4	70.0	256	21	AAH51601	Human prostate can
c	38	15.4	70.0	429	22	AAH51601	Probe #4676 for ge
c	39	15.4	70.0	429	22	AAH51601	Probe #4795 used t
c	40	15.4	70.0	429	22	AAH51601	Probe #4537 used t
	41	15.4	70.0	947	22	AAH51601	Nucleotide sequenc
	42	15.4	70.0	947	22	AAH51601	Nucleotide sequenc
	43	15.4	70.0	947	22	AAH51601	Apoptin-associatin
	44	15.4	70.0	1131	22	AAH51601	Nucleotide sequenc
	45	15.4	70.0	1131	22	AAH51601	Nucleotide sequenc

#### ALIGNMENTS

#### RESULT 1

AAH25458  
ID AAH25458 standard; DNA; 22 BP.

AC AAH25458;

DT 05-SEP-2001 (first entry)

DE PCR primer for endogenous sequences in transgenic plants.

KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.

OS Synthetic.

PN WO200141558-A1.

PD 14-JUN-2001.

PF 06-DEC-2000; 2000WO-EP12872.

PR 08-DEC-1999; 99US-0457037.

PA (AVET ) AVENTIS CROPS SCIENCE NV.

PI De Both G, De Beuckeleer M;

DR WPI; 2001-381419/40.

PT Transgenic winter oilseed rape plants suited for producing hybrid seed  
with improved qualities, comprises a male-sterility gene and fertility  
restorer gene, integrated into the genome

PS Example 5; Page 53; 98pp; English.

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAH5457-58 were used to amplify endogenous sequences  
 CC from transgenic plants of the invention.

XX Sequence 22 BP; 5 A; 5 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.023;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
 |||||  
 Db 1 cgcagttctgtgaacatcgacc 22

RESULT 2  
 AAD07003  
 ID AAD07003 standard; DNA; 22 BP.

AC AAD07003;  
 XX  
 DT 06-AUG-2001 (first entry)

XX PCR primer B04, to recognise foreign DNA and flanking sequence of MS-B2.  
 XX MS-B2 elite event; transgenic Brassica plant; transformation event;  
 XX male-sterility gene; PCR primer; ss.

XX Brassica napus.  
 XX WO200131042-A2.  
 XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-EF10680.

XX 29-OCT-1999; 99US-0430497.

XX (AVET ) AVENTIS CROPS SCIENCE NV.

XX Weston B, De Beuckeleer M;

XX WPI; 2001-300517/31.

XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome .

XX Example 5; Page 33; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.

XX The present sequence is PCR primer which is used to recognise foreign  
 CC DNA and a flanking sequence of elite event MS-B2.

XX Sequence 22 BP; 5 A; 7 C; 5 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.023;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
 |||||  
 Db 1 cgcagttctgtgaacatcgacc 22

RESULT 3  
 AAQ13870/C  
 ID AAQ13870 standard; DNA; 3113 BP.

XX AC AAQ13870;

XX 09-DEC-1991 (first entry)

DE Cruciferin A gene.

XX Seed storage protein; cruA; ss.

XX Brassica napus.

XX EP449376-A.

XX 02-OCT-1991.

XX 25-MAR-1991; 91EP-0200688.

XX 25-MAR-1991; 91EP-0200688.

XX 23-MAR-1990; 90US-0498561.

XX (KONN ) GIST-BROCADES NV.

XX Pen J, Sijmons PC, Van Ooyen AJJ, Rietveld K, Verwoerd TC;

XX Quax WJ;

XX WPI; 1991-289815/40.

XX Seeds contg. enhanced enzyme levels from transgenic plants - used  
 PT for catalysing reactions, increasing nutritional values or  
 PT treating digestive disorders.

XX Example; Fig 3; 38pp; English.

XX The DNA is the genomic sequence of the seed storage protein gene  
 CC cruciferin A (cruA). It can be used in the prodn. of transgenic  
 CC plants expressing cruciferin in its seeds for use in an industrial  
 CC process. The seeds contg. the cruciferin can be used without the  
 CC need for first extracting and/or isolating the enzymes. The use of  
 CC seeds for the storage of cruciferin provides a stable vehicle which  
 CC is easily packaged and transported and easily handled during use.  
 CC See also AAQ13871-Q13877.

XX Sequence 3113 BP; 961 A; 685 C; 586 G; 881 T; 0 other;

Query Match 100.0%; Score 22; DB 12; Length 3113;  
 Best Local Similarity 100.0%; Pred. No. 0.049;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
 |||||  
 Db 1183 CGCAGTTCGTGAACATCGACC 1162

RESULT 4  
 AAQ37508  
 ID AAQ37508 standard; DNA; 587 BP.

XX AC AAQ37508;

XX 17-JUN-1993 (first entry)

```

XX Shigella specific fragment NT15.
DE Chromosome; Shigella; sonnei; probe; Enteroinvasive E. coli; EIEC;
KW virulence plasmid; detection; dysentery; ss.
XX Shigella sonnei.
OS WO9303187-A.
PN 18-FEB-1993.
XX
XX 28-JUL-1992; 92WO-US06617.
PF
XX 31-JUL-1991; 91US-0738800.
PR
XX (STAD ) AMOCO CORP.
PA
XX McCarty JM, Parodos K;
PI WPI; 1993-076542/09.
DR
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
XX
XX Claim 3; Page 89-90; 129pp; English.
PS
XX The sequences given in AAQ37506-09 fragments which were derived from
CC the chromosome of Shigella sonnei. These fragments were used to
CC design probes which are specific to Shigella and Enteroinvasive E.
CC coli (EIEC). The probes are specific to a stable region of the
CC Shigella or E. coli genome, and not the unstable virulence plasmid,
CC which means that they are more reliable than previous probes in the
CC detection of dysentery causing microbes. See also AAQ37506-35.
XX
XX Sequence 587 BP; 138 A; 149 C; 139 G; 161 T; 0 other;
SQ
Query Match 78.2%; Score 17.2; DB 14; Length 587;
Best Local Similarity 86.4%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 cgcagttctgtgaacatcgacc 22
DB 566 cgcagtactgtgaacctcgatc 587
||||| |||||| |||||
RESULT 5
AAQ37510
ID AAQ37510 standard; DNA; 1174 BP.
XX
XX AAQ37510;
AC
XX 17-JUN-1993 (first entry)
DT
XX E. coli repeat 1.
DE
XX Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;
KW E. coli; transposable element; virulence plasmid; detection; dysentery;
KW repeat; ss.
XX
XX Escherichia coli.
OS
XX WO9303187-A.
PN
XX 18-FEB-1993.
PD
XX 28-JUL-1992; 92WO-US06617.
PF
XX 31-JUL-1991; 91US-0738800.
PR
XX (STAD ) AMOCO CORP.
PA
XX McCarty JM, Parodos K;
PI WPI; 1993-076542/09.
DR
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
XX

```

```

PA (STAD ) AMOCO CORP.
XX
XX McCarty JM, Parodos K;
DR WPI; 1993-076542/09.
XX
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
XX
XX Disclosure; Page 91; 129pp; English.
PS
XX The sequences given in AAQ37510-14 are fragments which represent repeat
CC sequences derived from the chromosome of E. coli and S. flexneri. The
CC repeat is highly conserved and has characteristics of a transposable
CC element. Over 20 copies of the repeat sequence are found in the
CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1
CC to 3 copies in some E. coli competitors, but not in other bacterial
CC species. These fragments were used to design probes which are
CC specific to Shigella and Enteroinvasive E. coli (EIEC). The probes
CC are specific to a stable region of the Shigella or E. coli genome, and
CC not the unstable virulence plasmid, which means that they are more
CC reliable than previous probes in the detection of dysentery causing
CC microbes. See also AAQ37506-35.
XX
XX Sequence 1174 BP; 252 A; 295 C; 277 G; 330 T; 20 other;
SQ
Query Match 78.2%; Score 17.2; DB 14; Length 1174;
Best Local Similarity 86.4%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 cgcagttctgtgaacatcgacc 22
DB 322 cgcagtactgtgaacctcgatc 343
||||| |||||| |||||
RESULT 6
AAQ37514
ID AAQ37514 standard; DNA; 1188 BP.
XX
XX AAQ37514;
AC
XX 17-JUN-1993 (first entry)
DT
XX S. flexneri repeat 2.
DE
XX Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;
KW E. coli; transposable element; virulence plasmid; detection; dysentery;
KW repeat; ss.
XX
XX Shigella flexneri.
OS
XX WO9303187-A.
PN
XX 18-FEB-1993.
PD
XX 28-JUL-1992; 92WO-US06617.
PF
XX 31-JUL-1991; 91US-0738800.
PR
XX (STAD ) AMOCO CORP.
PA
XX McCarty JM, Parodos K;
PI WPI; 1993-076542/09.
DR
XX Nucleic acid probes for detection of shigella and other pathogens
PT - used to diagnose dysentery in non-isotopic test format and have
PT utility in non-isotopic test formats requiring amplification for
PT high sensitivity
XX

```

PS Disclosure; Page 93; 129pp; English.

CC The sequences given in AAQ37510-14 are fragments which represent repeat  
 CC sequences derived from the chromosome of *E. coli* and *S. flexneri*. The  
 CC repeat is highly conserved and has characteristics of a transposable  
 CC element. Over 20 copies of the repeat sequence are found in the  
 CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1  
 CC to 3 copies in some *E. coli* competitors, but not in other bacterial  
 CC species. These fragments were used to design probes which are  
 CC specific to Shigella and Enteroinvasive *E. coli* (EIEC). The probes  
 CC are specific to a stable region of the Shigella or *E. coli* genome, and  
 CC not the unstable virulence plasmid, which means that they are more  
 CC reliable than previous probes in the detection of dysentery causing  
 CC microbes. See also AAQ37506-35.

XX Sequence 1188 BP; 258 A; 311 C; 273 G; 330 T; 16 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1188;  
 Best Local Similarity 86.4%; Pred. No. 15;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 ||||| ||||| ||||| |||||  
 Db 195 cgcagtactgtgaacctcgatc 216

RESULT 7

AAQ37512  
 ID AAQ37512 standard; DNA; 1196 BP.

AC AAQ37512;

DT 17-JUN-1993 (first entry)

DE *E. coli* repeat 2 (2).

KW Chromosome; Shigella; sonnei; flexneri; probe; Enteroinvasive; EIEC;  
 KW *E. coli*; transposable element; virulence plasmid; detection; dysentery;  
 KW repeat; ss.

OS *Escherichia coli*.

PN WO9303187-A.

PD 18-FEB-1993.

PF 28-JUL-1992; 92WO-US06617.

PR 31-JUL-1991; 91US-0738800.

PA (STAD ) AMOCO CORP.

XX McCarty JM, Parodos K;

XX WPI; 1993-076542/09.

PT Nucleic acid probes for detection of shigella and other pathogens  
 PT - used to diagnose dysentery in non-isotopic test format and have  
 PT utility in non-isotopic test formats requiring amplification for  
 PT high sensitivity

PS Disclosure; Page 92; 129pp; English.

CC The sequences given in AAQ37510-14 are fragments which represent repeat  
 CC sequences derived from the chromosome of *E. coli* and *S. flexneri*. The  
 CC repeat is highly conserved and has characteristics of a transposable  
 CC element. Over 20 copies of the repeat sequence are found in the  
 CC chromosome and virulence plasmid of Shigella. The repeat occurs in 1  
 CC to 3 copies in some *E. coli* competitors, but not in other bacterial  
 CC species. These fragments were used to design probes which are  
 CC specific to Shigella and Enteroinvasive *E. coli* (EIEC). The probes  
 CC are specific to a stable region of the Shigella or *E. coli* genome, and

CC not the unstable virulence plasmid, which means that they are more  
 CC reliable than previous probes in the detection of dysentery causing  
 CC microbes. See also AAQ37506-35.

XX Sequence 1196 BP; 259 A; 298 C; 272 G; 353 T; 14 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1196;  
 Best Local Similarity 86.4%; Pred. No. 15;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22  
 ||||| ||||| ||||| |||||  
 Db 195 cgcagtactgtgaacctcgatc 216

RESULT 8

AAV38672  
 ID AAV38672 standard; DNA; 2019 BP.

AC AAV38672;

DT 27-OCT-1998 (first entry)

DE *Mus musculus* SOCS7 gene.

KW SOCS; suppressor of cytokine signalling; PCR primer;  
 KW autoimmune disease; diagnosis; cancer; treatment;  
 KW cytokine mediated cellular responsiveness; hyperimmunity;  
 KW immunosuppression; allergies; hypertension; ss.

OS *Mus musculus*.

PH Key Location/Qualifiers  
 FT CDS 2..1054  
 FT /\*tag= a  
 FT /product= SOCS7 protein

PN WC9820023-A1.

PD 14-MAY-1998.

PF 31-OCT-1997; 97WO-AU00729.

PR 14-FEB-1997; 97AU-0005117.

PR 01-NOV-1996; 96AU-0003384.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Alexander WS, Hilton DJ, Metcalf D, Nicholson SE;

XX Nicola NA, Richardson RT, Starr R, Viney EM, Willson TA;

XX WPI; 1998-286854/25.

XX P-PSDB; AAW62621.

PT Suppressor of cytokine signalling proteins - useful to treat  
 PT disease, injury or abnormality involving cytokine mediated cellular  
 PT responsiveness e.g. hyperimmunity, immunosuppression, allergies and  
 PT hypertension

PS Claim 14; Page 143-144; 325pp; English.

XX The sequence is that of a gene encoding a suppressor of cytokine  
 CC signalling protein (SOCS). SOCS can be used to screen for naturally  
 CC occurring antibodies to SOCS, which may occur, e.g. in some autoimmune  
 CC diseases. Alternatively, specific antibodies can be used to  
 CC screen for SOCS, which is useful as a knowledge of SOCS levels  
 CC may be important for the diagnosis of certain cancers. Soluble  
 CC SOCS polypeptides can be used to treat disease, injury or  
 CC abnormality involving cytokine mediated cellular responsiveness,  
 CC e.g. hyperimmunity, immunosuppression, allergies and hypertension.

XX Sequence 2019 BP; 427 A; 510 C; 545 G; 528 T; 9 other;



Query Match 76.4%; Score 16.8; DB 19; Length 2019;  
 Best Local Similarity 90.0%; Pred. No. 26;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gcagttctgtgaacatcgac 21  
 |||||  
 Db 124 gcagttctgtgaccatccac 143

RESULT 9  
 AAH21115/c  
 ID AAH21115 standard; DNA; 7316 BP.  
 XX  
 AC AAH21115;  
 XX  
 DT 05-SEP-2001 (first entry)  
 XX  
 DE  
 XX C. glutamicum LP-6 DNA encoding tetr, teta and aada.  
 XX L-amino acid production; replication region; antibiotic resistance;  
 KW teta; tetracycline resistance; aada; streptomycin resistance; vitamin;  
 KW spectinomycin resistance; coryneform bacterium; D-panthothenic acid;  
 KW lysine; threonine; animal nutrition; food industry; medicine;  
 KW pharmaceutical industry; ds.  
 XX  
 OS Corynebacterium glutamicum.

FH Key Location/Qualifiers  
 FT CDS complement (1444..2013)  
 FT /\*tag= a  
 FT /product= "tetR"  
 FT 2124..3275  
 FT /\*tag= b  
 FT /product= "tetA"  
 FT 5882..6721  
 FT /\*tag= c  
 FT /product= "aada"  
 XX  
 PN EPI097998-A1.  
 XX  
 PD 09-MAY-2001.  
 XX  
 PF 11-OCT-2000; 2000EP-0122056.  
 XX  
 PR 05-NOV-1999; 99DE-1053206.  
 XX  
 PA (DEGS ) DEGUSSA AG.  
 XX  
 XX Tauch A, Kalinowski J, Puehler A, Thierbach G;  
 XX  
 XX WPI; 2001-391631/42.  
 DR  
 DR P-PSDB; AAB86252, AAB86253, AAB86254.  
 XX  
 XX New plasmids pTET3 and pCRY4, and their composites, useful for  
 PT expressing genes in coryneforms for production of amino acids, vitamins  
 PT and nucleotides .  
 XX  
 XX Claim 1 (1.3); Page 29-35; 46pp; German.  
 PS  
 XX This invention describes the novel plasmids pTET3 and pCRY4, isolated  
 CC from Corynebacterium glutamicum DSM 5816. pTET3 contains about 27.8 kb;  
 CC includes a replication region of 4539 bp (1) and a 7136 bp antibiotic  
 CC resistance region (6) containing the teta gene (resistance to  
 CC tetracycline) and the aada gene (resistance to streptomycin and  
 CC spectinomycin. pCRY4 contains about 48 kbp and includes a replication  
 CC region of 1856 bp (4). Restriction maps and sequences for (1), (4) and  
 CC (6) are reproduced. Composite plasmids derived from pTET3 and pCRY4 are  
 CC used to produce strains of coryneform bacteria that produce vitamins  
 CC (especially D-panthothenic acid); nucleotides, or L-amino acids  
 CC (particularly lysine and threonine), which are useful in animal  
 CC nutrition; the food and pharmaceutical industries, and human medicine.

CC Expression plasmids based on pTET3 or pCRY4 provide high productivity and  
 CC have unexpectedly good compatibility with known plasmids. This sequence  
 CC represents a DNA sequence containing a fragment which encodes the  
 CC Corynebacterium glutamicum LP-6 tetr, teta and aada proteins described in  
 CC the method of the invention.  
 XX  
 SQ Sequence 7316 BP; 1485 A; 2116 C; 2164 G; 1551 T; 0 other;

Query Match 76.4%; Score 16.8; DB 22; Length 7316;  
 Best Local Similarity 90.0%; Pred. No. 31;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20  
 |||||  
 Db 3778 CGCAGTTCTGCGATCATCGA 3759

RESULT 10  
 AAC77036/c  
 ID AAC77036 standard; cDNA; 4961 BP.  
 XX  
 AC AAC77036;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF2591 polynucleotide sequence SEQ ID NO:5181.  
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PF 31-MAR-1999; 99US-0127607.  
 PR  
 PR 02-APR-1999; 99US-0127636.  
 PR  
 PR 05-APR-1999; 99US-0127728.  
 PR  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Leach M;  
 PI  
 XX WPI; 2000-602362/57.  
 DR  
 DR P-PSDB; AAB42827.  
 XX  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 XX Claim 5; Page 4360-4363; 5507pp; English.  
 PS  
 XX AAC774446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 4961 BP; 1411 A; 1065 C; 1123 G; 1362 T; 0 other;

Query Match 74.5%; Score 16.4; DB 21; Length 4961;  
 Best Local Similarity 94.4%; Pred. No. 48;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatgcac 21  
 |||||  
 Db 2639 AGTCTGTGAACATTGAC 2622

RESULT 11  
 AAC12754/c  
 ID AAC12754 standard; cDNA; 226 BP.

XX AAC12754;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 16829.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Mline Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 16829; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX Sequence 226 BP; 75 A; 29 C; 51 G; 67 T; 4 other;

Query Match 71.8%; Score 15.8; DB 21; Length 226;  
 Best Local Similarity 89.5%; Pred. No. 62;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 cagttctgtgaacatgcac 21  
 |||||  
 Db 138 CAGTCTGTGAATTTCGAC 120

RESULT 12  
 AAH51601/c  
 ID AAH51601 standard; DNA; 319608 BP.

XX AC AAH51601;

XX 29-AUG-2001 (first entry)

XX Human chromosome 13q31-q33 genomic nucleotide sequence.

XX sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;  
 KW biallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.

XX Homo sapiens.

XX WO200058510-A2.

XX 05-OCT-2000.

XX 30-MAR-2000; 2000WO-IB00435.

XX 30-MAR-1999; 99US-0126903.

XX 30-APR-1999; 99US-0131971.

XX 30-APR-1999; 99US-0132065.

XX 14-JUL-1999; 99US-0143928.

XX 27-JUL-1999; 99US-0145915.

XX 29-JUL-1999; 99US-0146452.

XX 29-JUL-1999; 99US-0146453.

XX 28-OCT-1999; 99US-0162288.

XX (GEST ) GENSET.

XX Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihaun B;

XX Essioux L;

XX WPI; 2000-619082/59.

XX Polynucleotides comprising sequences from sbg1 and g35018 biallelic  
 PT markers are used for genotyping and detecting schizophrenia or bipolar  
 PT disorder and predisposition to these disorders -

XX Claim 1; Page 409-493; 737pp; English.

XX AAH51601 represents a human genomic nucleotide sequence comprising sbg1,  
 CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the  
 CC human chromosome 13q31-q33 locus. The nucleotide sequences contain  
 CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and  
 CC AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein  
 CC products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018  
 CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used  
 CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are  
 CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of  
 CC amplicons which comprise biallelic markers located on the chromosome  
 CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers

CC are represented in the sequences by degenerate/undefined base codes. PCR  
CC primers AAH51818 and AAH51819 are used in the isolation of sequences of  
CC the invention. The biallelic marker containing nucleotide sequences are  
CC used to determine the identity of the nucleotide at a biallelic marker in  
CC a sample DNA sequence. The nucleotide sequences may be labelled and used  
CC for genotyping by determining the identity of a nucleotide at a Region  
CC D-related biallelic marker in a biological sample from single or multiple  
CC subjects. By determining the frequency of a biallelic marker in a  
CC population an association between a genotype and a trait, a haplotype and  
CC a trait and a phenotype and a trait can be detected. The sequences can be  
CC used to determine a predisposition to or early onset of schizophrenia or  
CC bipolar disorder or a beneficial response to or side effects related to  
CC treatment against schizophrenia or bipolar disorder.  
XX

SQ Sequence 319608 BP; 101600 A; 56677 C; 58335 G; 102722 T; 274 other;  
Query Match 71.8%; Score 15.8; DB 21; Length 319608;  
Best Local Similarity 85.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gcagttctgtgaacatcga 20  
||||||| ||||||| |  
Db 164729 GCAGTTCTCTGACATCTA 164711

RESULT 13  
AAS09301/C  
ID AAS09301 standard; DNA; 319608 BP.

XX AC AAS09301;

XX DT 26-SEP-2001 (first entry)

XX DE Human schizophrenia associated gene g35030 and biallelic markers Al-A71.

XX KW Human; g35030; biallelic marker; Al-A71; chromosome 13q31-q33;  
XX RW schizophrenia; bipolar disorder; ds.

XX OS Homo sapiens.

EH Key Location/Qualifiers  
FH primer\_bind 7938..7958  
FT /\*tag= a  
FT /note= "Binds primer 99-27943.rp"  
FT primer\_bind 8297..8315  
FT /\*tag= b  
FT /note= "Binds primer 99-27943-150.mis"  
FT misc\_binding 8304..8328  
FT /\*tag= c  
FT /bound\_moiety= Probe\_99-27943-150  
FT misc\_feature 8316  
FT /\*tag= d  
FT /note= "Biallelic marker A1"  
FT primer\_bind complement (8317..8335)  
FT /\*tag= e  
FT /note= "Binds primer 99-27943-150.mis complement"  
FT primer\_bind complement (8446..8465)  
FT /\*tag= f  
FT /note= "Binds primer 99-27943.pu complement"  
FT primer\_bind 21365..21385  
FT /\*tag= g  
FT /note= "Binds primer 99-27935.rp"  
FT primer\_bind 21653..21671  
FT /\*tag= h  
FT /note= "Binds primer 99-27935-193.mis"  
FT misc\_binding 21660..21684  
FT /\*tag= i  
FT /bound\_moiety= Probe\_99-27935-193  
FT misc\_feature 21672  
FT /\*tag= j  
FT /note= "Biallelic marker A2"  
FT primer\_bind complement (21673..21691)

FT primer\_bind /\*tag= k  
FT /note= "Binds primer 99-27935-193.mis complement"  
FT complement (21845..21864)  
FT /\*tag= l  
FT /note= "Binds primer 99-27935.pu complement"  
FT primer\_bind 65463..65471  
FT /\*tag= m  
FT /note= "Binds primer 8-128.pu"  
FT primer\_bind 65466..65484  
FT /\*tag= n  
FT /note= "Binds primer 8-128-33.mis"  
FT misc\_binding 65473..65497  
FT /\*tag= o  
FT /bound\_moiety= Probe\_8-128-33  
FT misc\_feature 65485  
FT /\*tag= p  
FT /note= "Biallelic marker A3"  
FT primer\_bind complement (65486..65504)  
FT /\*tag= q  
FT /note= "Binds primer 8-128-33.mis complement"  
FT primer\_bind complement (65856..65874)  
FT /\*tag= r  
FT /note= "Binds primer 8-128.rp complement"  
FT primer\_bind 95034..95053  
FT /\*tag= s  
FT /note= "Binds primer 99-31960.pu"  
FT primer\_bind 95377..95395  
FT /\*tag= t  
FT /note= "Binds primer 99-31960-363.mis"  
FT misc\_binding 95384..95408  
FT /\*tag= u  
FT /bound\_moiety= Probe\_99-31960-363  
FT misc\_feature 95396  
FT /\*tag= v  
FT /note= "Biallelic marker A4"  
FT primer\_bind complement (95397..95415)  
FT /\*tag= w  
FT /note= "Binds primer 99-31960-363.mis complement"  
FT primer\_bind complement (95543..95563)  
FT /\*tag= x  
FT /note= "Binds primer 99-31960.rp complement"  
FT primer\_bind 107022..107040  
FT /\*tag= y  
FT /note= "Binds primer 99-24656.pu"  
FT primer\_bind 107262..107280  
FT /\*tag= z  
FT /note= "Binds primer 99-24656-260.mis"  
FT misc\_binding 107269..107293  
FT /\*tag= aa  
FT /bound\_moiety= Probe\_99-24656-260  
FT misc\_feature 107281  
FT /\*tag= ab  
FT /note= "Biallelic marker A5"  
FT primer\_bind complement (107282..107300)  
FT /\*tag= ac  
FT /note= "Binds primer 99-24656-260.mis complement"  
FT primer\_bind complement (107495..107513)  
FT /\*tag= ad  
FT /note= "Binds primer 99-24656.rp complement"  
FT primer\_bind 160279..160298  
FT /\*tag= ae  
FT /note= "Binds primer 99-24639.rp"  
FT primer\_bind 160621..160639  
FT /\*tag= af  
FT /note= "Binds primer 99-24639-163.mis"  
FT misc\_binding 160628..160652  
FT /\*tag= ag  
FT /bound\_moiety= Probe\_99-24639-163  
FT misc\_feature 160640  
FT /\*tag= ah  
FT /note= "Biallelic marker A6"  
FT primer\_bind complement (160641..160659)  
FT /\*tag= ai

```
FT primer_bind /note= "Binds primer 99-24639-163.mis complement"  
FT 160770..160787  
FT /tag= aj  
FT /note= "Binds primer 99-24634.pu"  
FT complement (160785..160802)  
FT /tag= ak  
FT /note= "Binds primer 99-24639.pu complement"  
FT 160857..160875  
FT /tag= al  
FT /note= "Binds primer 99-24634-108.mis"  
FT 160864..160888  
FT /tag= am  
FT /bound_moiety= Probe_99-24634-108  
FT 160876  
FT /tag= an  
FT /note= "Biallelic marker A7"  
FT complement (160877..160895)  
FT /tag= ao  
FT /note= "Binds primer 99-24634-108.mis complement"  
FT complement (161240..161257)  
FT /tag= ap  
FT /note= "Binds primer 99-24634.rp complement"  
FT 168813..168830  
FT /tag= aq  
FT /note= "Binds primer 99-7652.pu"  
FT 168955..168973  
FT /tag= ar  
FT /note= "Binds primer 99-7652-162.mis"  
FT 168962..168986  
FT /tag= as  
FT /bound_moiety= Probe_99-7652-162  
FT 168974  
FT /tag= at  
FT /note= "Biallelic marker A8"  
FT complement (168975..168993)  
FT /tag= au  
FT /note= "Binds primer 99-7652-162.mis complement"  
FT complement (169331..169351)  
FT /tag= av  
FT /note= "Binds primer 99-7652.rp complement"  
FT 170666..170686  
FT /tag= aw  
FT /note= "Binds primer 99-16100.pu"  
FT 170791..170809  
FT /tag= ax  
FT /note= "Binds primer 99-16100-147.mis"  
FT 170798..170822  
FT /tag= ay  
FT /bound_moiety= Probe_99-16100-147  
FT 170810  
FT /tag= az  
FT /note= "Biallelic marker A9"  
FT complement (170811..170829)  
FT /tag= ba  
FT /note= "Binds primer 99-16100-147.mis complement"  
FT complement (171153..171173)  
FT /tag= bb  
FT /note= "Binds primer 99-16100.rp complement"  
FT 173085..173085  
FT /tag= bc  
FT /note= "Binds primer 99-5862.rp"  
FT 173339..173357  
FT /tag= bd  
FT /note= "Binds primer 99-5862-167.mis"  
FT 173346..173370  
FT /tag= be  
FT /bound_moiety= Probe_99-5862-167  
FT 173358  
FT /tag= bf  
FT /note= "Biallelic marker A10"  
FT complement (173359..173377)  
FT /tag= bg  
FT /note= "Binds primer 99-5862-167.mis complement"
```

```
FT primer_bind complement (173495..173514)  
FT /tag= bh  
FT /note= "Binds primer 99-5862.pu complement"  
FT 189753..189771  
FT /tag= bi  
FT /note= "Binds primer 99-5919.pu"  
FT 189938..189956  
FT /tag= bj  
FT /note= "Binds primer 99-5919-215.mis"  
  
Query Match 71.8%; Score 15.8; DB 22; Length 319608;  
Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 gcagttctgtgaacatcga 20  
||||||| ||||| |  
Db 164729 GCAGTTCTCTGAACATCTA 164711  
  
RESULT 14  
AAX24719  
ID AAX24719 standard; cDNA; 545 BP.  
XX  
AC AAX24719;  
XX  
DT 21-JUN-1999 (first entry)  
XX  
XX Tobacco leaf polyphenol oxidase cDNA clone TOBPP06.  
DE  
XX Polyphenol oxidase; banana; tobacco; transgenic plant;  
KW vaccine; browning; spoilage; ds.  
XX  
OS Nicotiana tabacum.  
PN WO9853080-A1.  
XX  
PD 26-NOV-1998.  
XX  
PF 19-MAY-1998; 98WO-AU00362.  
PR 19-MAY-1997; 97AU-0006849.  
XX  
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.  
XX  
PI Robinson SP;  
XX  
DR WPI: 1999-070152/06.  
DR P-PSDB; AAW97990.  
XX  
XX Nucleic acid encoding polyphenol oxidase from banana, tobacco and  
PT pineapple - useful for, e.g. increasing levels of the enzyme  
PT expression, which is responsible for browning and spoilage of fruits  
PT after injury or damage  
XX  
PS Example 2; Fig 5; 47pp; English.  
XX  
CC This partial cDNA clone, termed TOBPP06, encodes a polyphenol oxidase  
CC (PPO) polypeptide of tobacco (see also AAW97990). The clone was  
CC obtained by PCR amplification of young tobacco leaf cDNA using  
CC primers (see AAX24708-12) based on conserved copper binding sites of  
CC plant PPOs, and identified on the basis of homology to known plant  
CC PPO genes. PPO is the major enzyme responsible for browning and  
CC spoilage of fruits and vegetables after damage or injury. Sense  
CC PPO nucleic acids can be used to increase the levels of PPO in a  
CC plant. Optionally modified sense sequences and antisense sequences  
CC can be used to reduce PPO levels (by co-suppression in the case of  
CC sense sequences). Transgenic plants including antisense PPO  
CC banana, tobacco or pineapple PPO or an antisense sequence are  
CC claimed.  
XX  
SQ Sequence 545 BP; 171 A; 115 C; 115 G; 144 T; 0 other;
```

```
Query Match      70.9%; Score 15.6; DB 20; Length 545;
Best Local Similarity 81.8%; Pred. No. 90;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22
   ||| ||||| ||| |||
Db 124 cgcggtctgtgaatcatggacc 145

RESULT 15
AAX30779
ID AAX30779 standard; DNA; 769 BP.
XX
AC AAX30779;
XX
XX 20-MAY-1999 (first entry)
XX
DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:56.
XX
KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;
XX streptococcal infection; pneumococcal; ss.
XX
OS Streptococcus pneumoniae.
XX
XX WO9737026-A1.
XX
XX 09-OCT-1997.
XX
XX 01-APR-1997; 97WO-US05306.
XX
XX 22-AUG-1996; 96US-0025788.
XX 02-APR-1996; 96US-0014690.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
XX Stodola RK;
XX
XX WPI; 1997-503111/46.
XX P-PSDB; AAY11181.
XX
XX Nucleic acids encoding pneumococcal polypeptide(s) - useful in
XX vaccines, drug screening, etc
XX
XX Claim 5; Page 91; 354pp; English.
XX
XX AAX30724 to AAX30946 represent genomic DNA sequences isolated from
XX Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
XX encode the novel proteins given in AAY11114 to AAY11367. The proteins,
XX isolated from Streptococcus pneumoniae, can be used in vaccines against
XX streptococcal infections and in assays for identifying compounds that
XX inhibit or activate the activity of the proteins. The antagonists can
XX be used to treat an individual having need to inhibit a bacterial
XX protein. Vectors expressing the proteins can be used to induce a
XX protective immune response in mammals.
XX
SQ Sequence 769 BP; 204 A; 202 C; 140 G; 223 T; 0 other;

Query Match      70.9%; Score 15.6; DB 18; Length 769;
Best Local Similarity 81.8%; Pred. No. 95;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgacc 22
   ||||| ||| |||
Db 5 cgcagttctgttaccacagacc 26
```

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:55 ; Search time 301.6 seconds  
(without alignments)  
16.520 Million cell updates/sec

Title: US-09-698-903B-14

Perfect score: 22  
Sequence: 1 cgcagttctgtgacatcgacc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/1na/PTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/1na/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	3113	1 US-08-146-422-20	Sequence 20, Appl
C 2	22	100.0	3113	1 US-08-628-534-2	Sequence 2, Appl
C 3	17.2	78.2	587	1 US-08-375-241-3	Sequence 3, Appl
C 4	17.2	78.2	587	5 PCT-US92-06617A-3	Sequence 3, Appl
C 5	17.2	78.2	1174	1 US-08-375-241-5	Sequence 5, Appl
C 6	17.2	78.2	1174	5 PCT-US92-06617A-5	Sequence 5, Appl
C 7	17.2	78.2	1188	1 US-08-375-241-9	Sequence 9, Appl
C 8	17.2	78.2	1188	5 PCT-US92-06617A-9	Sequence 9, Appl
C 9	17.2	78.2	1196	1 US-08-375-241-7	Sequence 7, Appl
C 10	17.2	78.2	1196	5 PCT-US92-06617A-7	Sequence 7, Appl
C 11	15.2	69.1	195	1 US-08-158-189-9	Sequence 9, Appl
C 12	15.2	69.1	1126	1 US-08-233-788A-48	Sequence 48, Appl
C 13	15.2	69.1	1576	3 US-08-689-974-2	Sequence 2, Appl
C 14	15.2	69.1	1576	3 US-09-058-376-2	Sequence 2, Appl
C 15	15.2	69.1	2048	1 US-07-602-608-11	Sequence 11, Appl
C 16	15.2	69.1	2048	1 US-08-261-578-11	Sequence 11, Appl
C 17	15.2	69.1	2158	1 US-07-602-608-1	Sequence 1, Appl
C 18	15.2	69.1	2158	1 US-08-261-578-1	Sequence 1, Appl
C 19	14.8	67.3	30	1 US-07-602-608-18	Sequence 18, Appl
C 20	14.8	67.3	30	1 US-08-261-578-18	Sequence 18, Appl
C 21	14.8	67.3	425	4 US-09-328-111-19	Sequence 19, Appl
C 22	14.6	66.4	952	4 US-09-174-768-2	Sequence 2, Appl
C 23	14.6	66.4	1062	2 US-08-468-819-92	Sequence 92, Appl
C 24	14.6	66.4	3330	1 US-08-149-103-1	Sequence 1, Appl
C 25	14.6	66.4	3330	1 US-08-451-883-1	Sequence 1, Appl
C 26	14.6	66.4	3656	1 US-08-393-734-1	Sequence 1, Appl
C 27	14.6	66.4	3656	4 US-08-894-489-1	Sequence 1, Appl

Query Match 100.0%; Score 22; DB 1; Length 3113;  
Best Local Similarity 100.0%; Pred. No. 0.012;

ALIGNMENTS

RESULT 1  
US-08-146-422-20/c  
; Sequence 20, Application US/08146422  
; Patent No. 5543576  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOLJEN, ALBERT J. J.  
; APPLICANT: RIETVELD, KRIJN  
; APPLICANT: HOEKEMA, ANDREAS  
; APPLICANT: PEN, JAN  
; APPLICANT: SIJMONS, PETER C.  
; APPLICANT: VERWOERD, TEUNIS C.  
; APPLICANT: QUAX, WILHEMUS J.  
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/146,422  
; FILING DATE: 02-NOV-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KENNEDY, BILL  
; REGISTRATION NUMBER: 33,407  
; REFERENCE/DOCKET NUMBER: 44615-20011.23  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3113 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-146-422-20

28	14.6	66.4	4649	6	5183745-1	Patent No. 5183745
29	14.6	66.4	6443	6	5183745-5	Patent No. 5183745
C 30	14.6	66.4	9592	1	US-08-393-734-3	Sequence 3, Appl
C 31	14.6	66.4	9592	4	US-08-894-489-3	Sequence 3, Appl
C 32	14.6	66.4	80161	3	US-09-036-987A-1	Sequence 1, Appl
C 33	14.6	66.4	80161	4	US-09-370-700-1	Sequence 1, Appl
C 34	14.4	65.5	1929	5	PCT-US93-00031-18	Sequence 18, Appl
35	14.4	65.5	1932	5	PCT-US93-00031-20	Sequence 20, Appl
36	14.4	65.5	1941	5	PCT-US93-00031-10	Sequence 10, Appl
37	14.4	65.5	1941	5	PCT-US93-00031-22	Sequence 22, Appl
38	14.4	65.5	2205	5	PCT-US93-00031-12	Sequence 12, Appl
39	14.4	65.5	2208	5	PCT-US93-00031-14	Sequence 14, Appl
40	14.4	65.5	2217	5	PCT-US93-00031-8	Sequence 8, Appl
41	14.4	65.5	2220	5	PCT-US93-00031-16	Sequence 16, Appl
42	14.4	65.5	2811	4	US-08-482-073-3	Sequence 3, Appl
43	14.4	65.5	2813	2	US-08-344-155C-99	Sequence 99, Appl
44	14.4	65.5	2813	4	US-09-009-490A-90	Sequence 90, Appl
45	14.4	65.5	3080	4	US-08-482-073-4	Sequence 4, Appl

```

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
    |||||
Db 1183 CGCAGTTCTGTGAACATCGACC 1162

RESULT 2
US-08-626-554-2/c
; Sequence 2, Application US/08626554
; Patent No. 5714474
; GENERAL INFORMATION:
; APPLICANT: VAN OOIJEN, ALBERT J.J.
; APPLICANT: RIETVELD, KRIJN
; APPLICANT: HOEKEMA, ANDREAS
; APPLICANT: PEN, JAN
; APPLICANT: SIJMONS, PETER C.
; APPLICANT: VERWOERD, TEUNIS C.
; APPLICANT: QUAX, WILHEMUS J.
; TITLE OF INVENTION: PRODUCTION OF ENZYMES IN SEEDS AND THEIR
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE NW
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/626,554
; FILING DATE: 02-APR-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 26192-20011.10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-626-554-2

Query Match 100.0%; Score 22; DB 1; Length 3113;
Best Local Similarity 100.0%; Pred. NO. 0.012;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
    |||||
Db 1183 CGCAGTTCTGTGAACATCGACC 1162

RESULT 3
US-08-375-241-3
; Sequence 3, Application US/08375241
; Patent No. 5648481
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarty, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella

```

```

; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,241
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: GTR90-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 587 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-375-241-3

Query Match 78.2%; Score 17.2; DB 1; Length 587;
Best Local Similarity 86.4%; Pred. NO. 3.1;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22
    |||||
Db 566 CGCAGTACTGTGAACCTCGATC 587

RESULT 4
PCT-US92-06617A-3
; Sequence 3, Application PC/TUS9206617A
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarty, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corporation
; STREET: 200 East Randolph Drive, P.O. Box 87703
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06617A
; FILING DATE: 19920728
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:

```



NAME: Galloway, Norval B.  
REGISTRATION NUMBER: 33,595  
REFERENCE/DOCKET NUMBER: GTR90-04 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-856-7180  
TELEFAX: 312-856-4972  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 587 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
PCT-US92-06617A-3

Query Match 78.2%; Score 17.2; DB 5; Length 587;  
Best Local Similarity 86.4%; Pred. No. 3.1;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||| ||||| |||||  
Db 566 CGCAGTACTGTGAACCTCGATC 587

RESULT 5  
US-08-375-241-5  
; Sequence 5, Application US/08375241  
; Patent No. 5648481  
; GENERAL INFORMATION:  
; APPLICANT: Parodos, Kyriaki  
; APPLICANT: McCarty, Janice  
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
; TITLE OF INVENTION: Shigella  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,241  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/738,800  
; FILING DATE: 31-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: GTR90-04  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1174 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-375-241-5

Query Match 78.2%; Score 17.2; DB 1; Length 1174;  
Best Local Similarity 86.4%; Pred. No. 3.6;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||| ||||| |||||  
Db 322 CGCAGTACTGTGAACCTCGATC 343

RESULT 6  
PCT-US92-06617A-5  
; Sequence 5, Application PC/TUS9206617A  
; GENERAL INFORMATION:  
; APPLICANT: Parodos, Kyriaki  
; APPLICANT: McCarty, Janice  
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
; TITLE OF INVENTION: Shigella  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amoco Corporation  
; STREET: 200 East Randolph Drive, P.O. Box 87703  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60680  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/06617A  
; FILING DATE: 19920728  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/738,800  
; FILING DATE: 31-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Galloway, Norval B.  
; REGISTRATION NUMBER: 33,595  
; REFERENCE/DOCKET NUMBER: GTR90-04 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-856-7180  
; TELEFAX: 312-856-4972  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1174 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
PCT-US92-06617A-5

Query Match 78.2%; Score 17.2; DB 5; Length 1174;  
Best Local Similarity 86.4%; Pred. No. 3.6;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgcagttctgtgaacatcgacc 22  
||||| ||||| ||||| |||||  
Db 322 CGCAGTACTGTGAACCTCGATC 343

RESULT 7  
US-08-375-241-9  
; Sequence 9, Application US/08375241  
; Patent No. 5648481  
; GENERAL INFORMATION:  
; APPLICANT: Parodos, Kyriaki  
; APPLICANT: McCarty, Janice  
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of  
; TITLE OF INVENTION: Shigella  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.

```
;
;
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,241
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: GTR90-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-375-241-9
```

```
Query Match 78.2% Score 17.2; DB 1; Length 1188;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 cgcagttctgtgaacatcgacc 22
||||| ||||||| |||||
Db 195 CGCAGTACTGTGAACCTCGATC 216
```

```
RESULT 8
PCT-US92-06617A-9
; Sequence 9, Application PC/TUS9206617A
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarty, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corporation
; STREET: 200 East Randolph Drive, P.O. Box 87703
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06617A
; FILING DATE: 19920728
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, Norval B.
; REGISTRATION NUMBER: 33,595
; REFERENCE/DOCKET NUMBER: GTR90-04 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-856-7180
; TELEFAX: 312-856-4972
; INFORMATION FOR SEQ ID NO: 9:
```

```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1188 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; PCT-US92-06617A-9
;
; Query Match 78.2% Score 17.2; DB 5; Length 1188;
; Best Local Similarity 86.4%; Pred. No. 3.6;
; Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
; Qy 1 cgcagttctgtgaacatcgacc 22
; ||||| ||||||| |||||
; Db 195 CGCAGTACTGTGAACCTCGATC 216
;
; RESULT 9
; US-08-375-241-7
; Sequence 7, Application US/08375241
; Patent No. 5648481
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarty, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,241
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: GTR90-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-375-241-7
```

```
Query Match 78.2% Score 17.2; DB 1; Length 1196;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 cgcagttctgtgaacatcgacc 22
||||| ||||||| |||||
Db 195 CGCAGTACTGTGAACCTCGATC 216
```

```
RESULT 10
PCT-US92-06617A-7
```

```

; Sequence 7, Application PC/TUS9206617A
; GENERAL INFORMATION:
; APPLICANT: Parodos, Kyriaki
; APPLICANT: McCarthy, Janice
; TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
; TITLE OF INVENTION: Shigella
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amoco Corporation
; STREET: 200 East Randolph Drive, P.O. Box 87703
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06617A
; FILING DATE: 19920728
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/738,800
; FILING DATE: 31-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Galloway, Norval B.
; REGISTRATION NUMBER: 33,595
; REFERENCE/DOCKET NUMBER: GTR90-04 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-856-7180
; TELEFAX: 312-856-4972
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1196 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US92-06617A-7

```

```

Query Match 78.2%; Score 17.2; DB 5; Length 1196;
Best Local Similarity 86.4%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatgcacc 22
||||| ||||||| |||||
Db 195 CGCAGTACTGTGACCTCGATC 216

RESULT 11
US-08-158-189-9
; Sequence 9, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides.
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/158,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Phillip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-158-189-9

Query Match 69.1%; Score 15.2; DB 1; Length 195;
Best Local Similarity 85.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20
||||| ||||| |||||
Db 130 CACAGTTCAGTGAGCATCGA 149

RESULT 12
US-08-233-788A-48/c
; Sequence 48, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1126 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

RESULT 14  
US-09-058-376-2/c  
; Sequence 2, Application US/09058376

RESULT 15  
US-07-602-608-11/c  
; Sequence 11, Application US/07602608  
; Patent No. 5382524  
; GENERAL INFORMATION:  
; APPLICANT: Desnick, Robert J.  
; APPLICANT: Bishop, David F.  
; APPLICANT: Ioannou, Yiannis A.  
; APPLICANT: Wang, Anne M.  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY  
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/602,608  
; FILING DATE: 24-OCT-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 6923-008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2048 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; US-07-602-608-11

Query Match 69.1%; Score 15.2; DB 1; Length 2048;  
Best Local Similarity 85.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcga 20  
||||| ||||| |||||  
Db 101 CGCAGGTCGTGGACATCAA 82

Search completed: February 25, 2002, 18:05:56  
Job time: 18604 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:21 : Search time 8261.74 Seconds  
(without alignments)  
28.615 Million cell updates/sec

Title: US-09-698-903B-14  
Perfect score: 22  
Sequence: 1 cgcagttctgtgaacatgacc 22  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_estfun:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estom:*
5:	em_estpl:*
6:	em_estba:*
7:	em_estro:*
8:	em_estov:*
9:	em_htc:*
10:	gb_est1:*
11:	gb_est2:*
12:	gb_htc:*
13:	gb_gss:*
14:	em_gss_fun:*
15:	em_gss_hum:*
16:	em_gss_inv:*
17:	em_gss_pln:*
18:	em_gss_pro:*
19:	em_gss_rod:*
20:	em_gss_vrt:*
21:	em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	80.9	653	10	BE377681 601229824
2	16.8	76.4	244	10	AI561687 vw88f09.x
3	16.8	76.4	279	13	AZ777284 2M0011K06
4	16.8	76.4	346	10	AI019782 ua90h12.r
5	16.8	76.4	605	13	AL026432 Fugu rubr
6	16.8	76.4	611	13	AL026427 Fugu rubr
7	16.8	76.4	613	13	AZ777336 2M0011I08
8	16.8	76.4	1301	12	AK006515 Mus muscu
9	16.4	74.5	206	10	AW086177 xc77ell.x
10	16.4	74.5	234	10	AA701011 zg55d04.s
11	16.4	74.5	270	11	F02043 HSCOMG102 n
12	16.4	74.5	290	10	AA569516 nf23e01.s

13	16.4	74.5	303	11	F03921
14	16.4	74.5	304	10	AA843570
15	16.4	74.5	335	10	AI084952
16	16.4	74.5	399	10	AI658711
17	16.4	74.5	412	11	BF066458
18	16.4	74.5	435	11	N49244
19	16.4	74.5	441	10	AI196345
20	16.4	74.5	455	11	BF483372
21	16.4	74.5	460	10	AA679411
22	16.4	74.5	463	10	AA426087
23	16.4	74.5	483	10	AI149964
24	16.4	74.5	487	10	AI802194
25	16.4	74.5	499	10	AA421463
26	16.4	74.5	508	10	AA700220
27	16.4	74.5	526	10	AA235369
28	16.4	74.5	545	11	R60174
29	16.4	74.5	553	10	AI631483
30	16.4	74.5	560	10	AW013257
31	16.4	74.5	582	11	W91993
32	16.4	74.5	612	13	BH121721
33	16.4	74.5	630	10	AA700001
34	16.4	74.5	650	10	AA952652
35	16.4	74.5	673	10	AA868429
36	16.4	74.5	679	10	AI680857
37	16.4	74.5	706	13	AZ717353
38	16.4	74.5	758	10	AI740628
39	16.4	74.5	768	10	AI637690
40	16.4	74.5	804	10	AL041092
41	16.4	74.5	833	10	BE412370
42	16.2	73.6	122	11	BG989620
43	16.2	73.6	400	13	AQ852470
44	16.2	73.6	406	13	AZ160523
45	16.2	73.6	431	10	AV655742

ALIGNMENTS

RESULT 1

BE377681 601229824F1 NCI\_CGAP\_Mam1 Mus musculus cdna clone IMAGE:3594076 5',  
LOCUS mRNA EST 21-JUL-2000  
DEFINITION mRNA sequence.  
ACCESSION BE377681  
VERSION BE377681.1 GI:9323046  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 653)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LIA88767 row: o column: 05  
High quality sequence stop: 612.  
Location/Qualifiers  
1. .653  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3594076"  
/clone\_lib="NCI\_CGAP\_Mam1"  
/tissue\_type="tumor, biopsy sample"

FEATURES

source

/dev\_stage="3 months, virgin"  
/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 132 a 181 c 195 g 145 t  
ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 653;  
Best Local Similarity 90.5%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cgcagttctgtgaacatcgac 21  
||||| ||||| |||||  
Db 193 GCCAGTTCGTGAGCTCGAC 213

## RESULT 2

AI561687 244 bp mRNA EST 25-MAR-1999  
LOCUS  
DEFINITION vv88f09.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone  
IMAGE:1262057 3', mRNA sequence.

ACCESSION AI561687  
VERSION  
KEYWORDS  
SOURCE EST.  
GI:4513032

house mouse.  
Mus musculus

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 244)

REFERENCE  
AUTHORS  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson R.  
Waterston,R. and Wilson R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:664609

This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end

High quality sequence stop: 229.

Location/Qualifiers

1..244

/organism="Mus musculus"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:1262057"

/clone\_lib="Stratagene mouse skin (#937313)"

/sex="females"

/tissue\_type="whole skin"

/dev\_stage="11 weeks old"

/lab\_host="SOLR (kanamycin resistant)"

/note="Organ: skin; Vector: pBluescript SK-; Site\_1: EcoRI  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dt. Whole skin from 11 week old C57BL/6 female mice.

Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'

adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

53 a 53 c 51 g 87 t

BASE COUNT  
ORIGIN

Query Match 76.4%; Score 16.8; DB 10; Length 244;

Best Local Similarity 90.0%; Pred. No. 5.3e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 cagttctgtgaacatcgacc 22

||||| ||||| ||||| |||||

Db 85 CAGTTCCTTTGAACATCTACC 104

## RESULT 3

AZ777264

LOCUS

DEFINITION AZ777264 279 bp DNA GSS 16-FEB-2001  
clone UUGC2M0011K06 R, DNA sequence.

ACCESSION AZ777264

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 279)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.

and Wright,D. Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0011 row: K column: 06

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 279.

Location/Qualifiers

1..279

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0011K06"

/clone\_lib="Mouse 10kb plasmid UUGCLM library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/).

The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT  
ORIGIN

80 a 65 c 69 g 65 t



Query Match	76.4%; Score 16.8; DB 13; Length 279;
Best Local Similarity	90.0%; Pred. No. 5.5e+02;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 cgcagttctgtgaacatcga 20 
Db	162 CTCAGTTCTGTGACCATCCGA 181 
RESULT	4
LOCUS	A1019782
DEFINITION	us90h12.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone EST 16-JUN-1998
ACCESSION	IMAGE:I1364807 5', mRNA sequence.
VERSION	A1019782
KEYWORDS	A1019782.1 GI:3234118
SOURCE	EST.
ORGANISM	house mouse.
REFERENCE	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 346)
TITLE	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,F., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Stepcoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie.T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
JOURNAL	The WashU-HMI Mouse EST Project
COMMENT	Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mousee@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:898027
FEATURES	Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 285. Location/Qualifiers 1..346 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:I1364807" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B"
BASE COUNT	74 a 95 c 120 g 57 t
ORIGIN	
Query Match	76.4%; Score 16.8; DB 10; Length 346;
Best Local Similarity	90.0%; Pred. No. 5.7e+02;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	2 gcagttctgtgaacatcgac 21 

Query Match	76.4%; Score 16.8; DB 13; Length 279;
Best Local Similarity	90.0%; Pred. No. 5.5e+02;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 cgcagttctgtgaacatcga 20 
Db	162 CTCAGTTCTGTGACCATCCGA 181 
RESULT	4
LOCUS	A1019782
DEFINITION	us90h12.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone EST 16-JUN-1998
ACCESSION	IMAGE:I1364807 5', mRNA sequence.
VERSION	A1019782
KEYWORDS	A1019782.1 GI:3234118
SOURCE	EST.
ORGANISM	house mouse.
REFERENCE	Mus musculus
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 346)
TITLE	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,F., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Stepcoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie.T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
JOURNAL	The WashU-HMI Mouse EST Project
COMMENT	Unpublished (1996) Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mousee@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:898027
FEATURES	Seq primer: -28ml3 rev2 ET from Amersham High quality sequence stop: 285. Location/Qualifiers 1..346 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="IMAGE:I1364807" /sex="male" /tissue_type="mammary gland" /dev_stage="4 weeks" /lab_host="DH10B"
BASE COUNT	74 a 95 c 120 g 57 t
ORIGIN	
Query Match	76.4%; Score 16.8; DB 10; Length 346;
Best Local Similarity	90.0%; Pred. No. 5.7e+02;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	2 gcagttctgtgaacatcgac 21 

Query Match	76.4%; Score 16.8; DB 13; Length 605;
Best Local Similarity	90.0%; Pred. No. 6.3e+02;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 cgcagttctgtgaacatcga 20 
Db	29 CCCAGTTTCTGTGACATCGA 48 
RESULT	6
LOCUS	FR0030058
DEFINITION	Fugu rubripes GSS sequence, clone 072H16AG6, genomic survey sequence.
ACCESSION	AL026427
VERSION	AL026427.1 GI:3263770
KEYWORDS	GSS: genome survey sequence.
SOURCE	Takifugu rubripes
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu. 1 (bases 1 to 611)
REFERENCE	Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y., Williams,G. and Brenner,S.
TITLE	Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK. Email: biohelp@hmp.mrc.ac.uk
JOURNAL	Vector: pBluescript II KS
COMMENT	V_type: phagemid PRIMER: KS
BASE COUNT	139 a 147 c 173 g 14 others
ORIGIN	
Query Match	76.4%; Score 16.8; DB 13; Length 605;
Best Local Similarity	90.0%; Pred. No. 6.3e+02;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 cgcagttctgtgaacatcga 20 
Db	29 CCCAGTTTCTGTGACATCGA 48 
RESULT	6
LOCUS	FR0030058
DEFINITION	Fugu rubripes GSS sequence, clone 072H16AG6, genomic survey sequence.
ACCESSION	AL026427
VERSION	AL026427.1 GI:3263770
KEYWORDS	GSS: genome survey sequence.
SOURCE	Takifugu rubripes
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu. 1 (bases 1 to 611)
REFERENCE	Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y., Williams,G. and Brenner,S.
TITLE	Submitted (09-JUN-1998) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK. Email: biohelp@hmp.mrc.ac.uk
JOURNAL	Vector: pBluescript II KS
COMMENT	V_type: phagemid PRIMER: KS
BASE COUNT	139 a 147 c 173 g 14 others
ORIGIN	
Query Match	76.4%; Score 16.8; DB 13; Length 605;
Best Local Similarity	90.0%; Pred. No. 6.3e+02;
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy	1 cgcagttctgtgaacatcga 20 
Db	29 CCCAGTTTCTGTGACATCGA 48 
RESULT	6
LOCUS	FR0030058
DEFINITION	Fugu rubripes GSS sequence, clone 072H16AG6, genomic survey sequence.
ACCESSION	AL026427
VERSION	AL026427.1 GI:3263770
KEYWORDS	GSS: genome survey sequence.
SOURCE	Takifugu rubripes
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu. 1 (bases 1 to 611)
REFERENCE	Elgar,G., Clark,M., Smith,S., Meek,S





This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham.

## FEATURES

## source

```

Location/Qualifiers
1..234
/organism="Homo sapiens"
/db_xref="GDB:1303157"
/db_xref="taxon:9606"
/clone="IMAGE:397255"
/clone_lib="Soares_pineal_gland_N3HPC"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: pineal gland; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5', TGTTACCAATCTGAAGTGGAGCGCGCTTTTCTTTTCTTTT 3']
, double-stranded cDNA was size selected, ligated to Eco
RI adapters (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo."
79 a 31 c 35 g 89 t
BASE COUNT
ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 234;
Best Local Similarity 94.4%; Pred. No. 8.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21
|||||
Db 21 AGTTCTGTGAACATTTGAC 38

RESULT 11
LOCUS F02043
DEFINITION HSCOMG102 normalized infant brain cDNA Homo sapiens cDNA clone
c-0mg10 3', mRNA sequence.
ACCESSION F02043
VERSION F02043.1 GI:645600
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 270)
AUTHORS Auffray,C., Behar,G., Bols,F., Bouchier,C., da Silva,C., Devignes
,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchi,S. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read. removed at sequence 5'end
Genexpress_library_idt: C; Genexpress_sequence_idt: alc-0mg10
Seq primer: (-21)ml3-universal.
Location/Qualifiers
1..270
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-0mg10"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"

```

/note="Organ: brain; Vector: lafmid BA; Site\_1: HindIII; Site\_2: NotI; sex=Female; dev\_stage=3 months old; isolate=muscular atrophy patient; tissue\_type=total brain ; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Soaers, Psychiatry Dept. Columbia University, USA. Normalization\_method: Bento Soares, P.N.A.S in press"

BASE COUNT  
ORIGIN

Query Match 74.5%; Score 16.4; DB 11; Length 270;  
Best Local Similarity 94.4%; Pred. No. 8.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21

|||||

Db 23 AGTTCTGTGAACATTTGAC 40

## RESULT 12

## AA569516

## LOCUS

DEFINITION nf23e01.s1 NCI-CGAP\_Prl Homo sapiens cDNA clone IMAGE:914616, mRNA sequence.

## ACCESSION

## AA569516

## VERSION

## AA569516.1

## GI:2343496

## KEYWORDS

## EST.

## SOURCE

## human.

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## 1 (bases 1 to 290)

## AUTHORS

## National Cancer Institute,

## Tumor Gene Index

## Unpublished (1997)

## COMMENT

## Contact: Robert Strausberg, Ph.D.

## Email: cgapbs-r@mail.nih.gov

## Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.

## , Michael Emmert-Buck, M.D., Ph.D.

## cDNA Library Preparation: David B. Krizman, Ph.D.

## DNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.

## DNA Sequencing by: Washington University Genome Sequencing Center

## Clone distribution: NCI-CGAP clone distribution information can be

## found through the I.M.A.G.E. Consortium/LLNL at:

## www-bio.llnl.gov/bbrp/image/image.html

## Seq primer: -40ml3 fwd. ET from Amersham

## High quality sequence stop: 263.

## Location/Qualifiers

## 1..290

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## /clone="IMAGE:914616"

## /clone\_lib="NCI-CGAP\_Prl"

## /sex="Male"

## /dev\_stage="45 years old"

## /lab\_host="DH10B"

## /note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st

## strand cDNA was primed with oligo(dT)17 on 50 ng of

## DNase-treated, total cellular RNA obtained from 5,000-10

## ,000 microdissected, histologically normal prostate

## epithelial cells. Double-stranded cDNA was ligated to

## EcoRI adaptors, 5 cycles of PCR applied to the cDNA with

## an adaptor-specific primer, and the resulting PCR product

## subcloned into pAMP10 by the UDG-cloning method (Life

## Technologies). Average insert size is 600 bp. NOTE: Not

## directionally cloned. This library was constructed by

## David Krizman."

## 77 a 76 c 73 g 64 t

## BASE COUNT

## ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 290;  
 Best Local Similarity 94.4%; Pred. No. 8.6e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 gttctgtgaacatcgacc 22  
 ||||| ||||| ||||| |||||  
 Db 233 GTTCTGTGGACATCGACC 250

RESULT 13  
 F03921

LOCUS  
 DEFINITION HSC2DF112 normalized infant brain cDNA Homo sapiens CDNA clone  
 C-2df11 3', mRNA sequence.

ACCESSION F03921  
 VERSION F03921  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 303)  
 AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Degniges  
 , M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F.,  
 Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,  
 Sebastiani-Kabatchis, C. and Tessier, A.  
 IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 95277534

MEDLINE  
 COMMENT Contact: Genethon  
 Genexpress-Genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698  
 Email: genexpress@genethon.fr  
 Single read. removed at sequence 5' end  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-2df11  
 Seq primer: (-21)M13\_universal.

FEATURES  
 source Location/Qualifiers  
 1. .303  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="c-2df11"  
 /clone\_lib="normalized infant brain cDNA"  
 /sex="Female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: lafmid BA; Site\_1: HindIII;  
 Site\_2: NotI; sex=Female; dev\_stage=3 months old;  
 isolate=muscular atrophy patient; tissue\_type=total brain  
 ; total mRNA was oligo-(dT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lafmid BA vector. Clone library from B.Souares, Psychiatry  
 Dept. Columbia University, USA. Normalization\_method:  
 Bento Soares, P.N.A.S in press"  
 99 a 42 c 43 g 116 t 3 others

BASE COUNT  
 ORIGIN

Query Match 74.5%; Score 16.4; DB 11; Length 303;  
 Best Local Similarity 94.4%; Pred. No. 8.6e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21  
 ||||| ||||| ||||| |||||  
 Db 23 AGTCTGTGAACATTGAC 40

RESULT 14  
 AA843570

LOCUS  
 DEFINITION aJ54f09.s1 Soares\_testis\_NHT Homo sapiens CDNA clone IMAGE:1394153  
 3', mRNA sequence.

ACCESSION AA843570  
 VERSION AA843570.1  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 304)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb-remail.nih.gov  
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
 , Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 DNA distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 1108 Std Error: 0.00  
 Seq primer: -40ml3 fwd. EF from Amersham  
 High quality sequence stop: 284.

FEATURES  
 source Location/Qualifiers  
 1. .304  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1394153"  
 /clone\_lib="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech Laboratories  
 , Inc., and primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGGCGCGCCCAATTTTTTTTTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization to Cots5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo. "  
 99 a 43 c 43 g 119 t

BASE COUNT  
 ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 304;  
 Best Local Similarity 94.4%; Pred. No. 8.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 agttctgtgaacatcgac 21  
 ||||| ||||| ||||| |||||  
 Db 26 AGTCTGTGAACATTGAC 43

RESULT 15  
 AI084952

LOCUS  
 DEFINITION ow86c05.s1 Soares\_fetal\_liver\_spleen\_lNFLS\_s1 Homo sapiens CDNA  
 clone IMAGE:1653704 3', mRNA sequence.

ACCESSION AI084952  
 VERSION AI084952.1  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 335)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL  
COMMENT

## FEATURES

BASE COUNT  
ORIGIN

Search completed: February 25, 2002, 17:21:24  
Job time: 16177 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:53:12 ; Search time 2331.3 Seconds  
(without alignments)  
148.604 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21

Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	AX127749 Sequence
c 2	21	100.0	249	6	AR110601 Sequence
c 3	21	100.0	249	6	AR151000 Sequence
c 4	21	100.0	249	6	AR152423 Sequence
c 5	21	100.0	249	6	AR152432 Sequence
c 6	21	100.0	261	6	AR110597 Sequence
c 7	21	100.0	261	6	AR150996 Sequence
c 8	21	100.0	261	6	AR152419 Sequence
c 9	21	100.0	261	6	AR152428 Sequence
c 10	21	100.0	280	6	E01312
c 11	21	100.0	288	11	RICLA29B
c 12	21	100.0	318	1	ABA7623
c 13	21	100.0	358	1	ABA7624
c 14	21	100.0	563	6	A71436 Sequence
c 15	21	100.0	838	6	AX014764 Sequence
c 16	21	100.0	900	6	AX019188 Sequence
c 17	21	100.0	900	6	AX019212 Sequence
c 18	21	100.0	1036	6	AX014765 Sequence
c 19	21	100.0	1045	12	SCO308515
c 20	21	100.0	1138	6	A59870
c 21	21	100.0	1138	6	A94795 Sequence
c 22	21	100.0	1287	6	A18052
c 23	21	100.0	1287	6	AR095108 Sequence
c 24	21	100.0	1287	6	AR098314 Sequence
c 25	21	100.0	1287	6	AX012339 Sequence
c 26	21	100.0	1287	6	I49887 Sequence
c 27	21	100.0	1287	6	I82375 Sequence
c 28	21	100.0	1559	6	A84754 Sequence
c 29	21	100.0	1559	6	AR118432 Sequence
c 30	21	100.0	1690	6	AX021176 Sequence
c 31	21	100.0	1829	6	A26453 Recombinant
c 32	21	100.0	1829	6	I16753 Sequence
c 33	21	100.0	1863	6	A32909 Chimeric en
c 34	21	100.0	1863	6	A40108 Sequence
c 35	21	100.0	1863	6	AR022666 Sequence
c 36	21	100.0	1863	6	AR029592 Sequence
c 37	21	100.0	1949	6	A83076 Sequence
c 38	21	100.0	2020	8	D88395 Oryza sativ
c 39	21	100.0	2042	12	AF190131 Cloning v
c 40	21	100.0	2115	6	AR027732 Sequence
c 41	21	100.0	2145	6	AR019221 Sequence
c 42	21	100.0	2256	8	AB003140 Arabidops
c 43	21	100.0	2319	6	A84753 Sequence
c 44	21	100.0	2319	6	AR118431 Sequence
c 45	21	100.0	2345	6	AR118422 Sequence

ALIGNMENTS

RESULT 1	AX127749	AX127749	21 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 2 from Patent WO0131042.					
DEFINITION	AX127749					
ACCESSION	AX127749					
VERSION	AX127749.1	GI:14134396				
KEYWORDS	synthetic construct.					
SOURCE	synthetic construct					
ORGANISM	artificial sequence.					
REFERENCE	1 (bases 1 to 21)					
AUTHORS	Weston,B. and de Beuckeleer,M.					
TITLE	Male-sterile brassica plants and methods for producing same					
JOURNAL	Patent: WO 0131042-A 2 03-MAY-2001;					
FEATURES	Aventis CropScience N.V. (BE)					
source	Location/Qualifiers					
	1. .21					
	/organism="synthetic construct"					
	/db_xref="taxon:32630"					
	/note="primer MDB355"					

BASE COUNT 7 a 6 c 5 g 3 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||

Db 1 GTAACATAGATGACACCGCGC 21

## RESULT 2

AR110601/c  
LOCUS AR110601 249 bp DNA PAT 14-FEB-2001  
DEFINITION Sequence 8 from patent US 6114608.  
ACCESSION AR110601  
VERSION AR110601.1 GI:12826877  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Dietrich,P.S and Sinibaldi,R.M.  
TITLE Nucleic acid construct comprising bacillus thuringiensis cryIAb gene  
JOURNAL Patent: US 6114608-A 8 05-SEP-2000;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 3

AR151000/c  
LOCUS AR151000 249 bp DNA PAT 08-AUG-2001  
DEFINITION Sequence 8 from patent US 6229075.  
ACCESSION AR151000  
VERSION AR151000.1 GI:15115591  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and Gardiner,M.  
TITLE Inbred maize line R412H  
JOURNAL Patent: US 6229075-A 8 08-MAY-2001;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 4

AR152423/c  
LOCUS AR152423 249 bp DNA PAT 08-AUG-2001  
DEFINITION Sequence 8 from patent US 6232533.  
ACCESSION AR152423  
VERSION AR152423.1 GI:15118473  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and Gardiner,M.  
TITLE Inbred maize line R372H  
JOURNAL Patent: US 6232533-A 8 15-MAY-2001;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 5

AR152432/c  
LOCUS AR152432 249 bp DNA PAT 08-AUG-2001  
DEFINITION Sequence 8 from patent US 6232534.  
ACCESSION AR152432  
VERSION AR152432.1 GI:15118482  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 249)  
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and Gardiner,M.  
TITLE Inbred maize line R660H  
JOURNAL Patent: US 6232534-A 8 15-MAY-2001;  
FEATURES Location/Qualifiers  
source 1..249

BASE COUNT 79 a 38 c 47 g 85 t  
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21  
|||||

Db 247 GTAACATAGATGACACCGCGC 227

## RESULT 6

AR110597/c  
LOCUS AR110597 261 bp DNA PAT 14-FEB-2001  
DEFINITION Sequence 4 from patent US 6114608.  
ACCESSION AR110597  
VERSION AR110597.1 GI:12826873  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.



```
Unclassified.
REFERENCE 1 (bases 1 to 261)
AUTHORS Mettler,I.J., Dietrich,P.S and Sinibaldi,R.M.
TITLE Nucleic acid construct comprising bacillus thuringiensis cryIab
gene
JOURNAL Patent: US 6114608-A 4 05-SEP-2000;
FEATURES Location/Qualifiers
source 1..261
BASE COUNT 83 a 41 c 50 g 87 t
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
|||||
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 9
ARI52428/c
LOCUS ARI52428 261 bp DNA 08-AUG-2001
DEFINITION Sequence 4 from patent US 6232534.
ACCESSION ARI52428
VERSION ARI52428.1 GI:15118478
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 261)
AUTHORS Mettler,I.J., Plaisted,D.C., Grier,S.L., Houghton,W. and
Gardiner,W.
TITLE Inbred maize line R660H
JOURNAL Patent: US 6232534-A 4 15-MAY-2001;
FEATURES Location/Qualifiers
source 1..261
BASE COUNT 83 a 41 c 50 g 87 t
ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
|||||
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 10
E01312/c
LOCUS E01312 280 bp DNA 29-SEP-1997
DEFINITION 3' Untranslated region of nopalín synthetase gene.
ACCESSION E01312
VERSION E01312.1 GI:2169571
KEYWORDS JP 1987201527-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 280)
AUTHORS Rojlyaa,E.B., Robaato,T.F. and Suteilibun,J.R.
TITLE PROTECTION OF PLANT FROM VIRUS INFECTION
JOURNAL Patent: JP 1987201527-A 4 05-SEP-1987;
COMMENT MONSANTO CO, WASHINGTON UNIV
OS Agrobacterium
PN JP 1987201527-A/4
PD 05-SEP-1987
PF 29-OCT-1986 JP 1986258063
PR 29-OCT-1985 US 85 792389, 27-MAR-1986 US 86 844918, PR
09-OCT-1986 US 86 917027
PI ROJIVAA ENU BIICHI, ROBAATO TEII FURAREI,
PI SUTEILIBUN JII ROJIVAA SU
PC AO1H5/00,A01H1/00,C12N1/20,C12N15/00,C07K13/00, PC
(C12N1/20,
PC C12R1:01);
CC strandedness: Single;
CC topology: linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
```

```

FH
FT polyA_signal >1..<281.
FEATURES
  source
    Location/Qualifiers
      1..280
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT      83 a 49 c 58 g 90 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 280;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21
    |||||||
Db 246 GTAACATAGATGACACCGCGC 226

RESULT 11
R1C1429B/c
LOCUS
DEFINITION Rice genomic DNA, L429B, sequence tagged site.
ACCESSION D25449
VERSION D25449.1 GI:436695
KEYWORDS STS; Not I linking clone; RFLP; Rice.
SOURCE Oryza sativa (strain:Nipponbare) DNA.
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
  1 (bases 1 to 288)
  Minobe, Y.
  Direct Submission
  Submitted (18-NOV-1993) to the DDBJ/EMBL/GenBank databases. Yuzo
  Minobe, National Institute of Agrobiological Resources, Rice Genome
  Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305, Japan
  (E-mail:MINOBE@rsg.riken.go.jp, Tel:0298-38-7441,
  Fax:0298-38-7468)
REFERENCE
  2 (bases 1 to 288)
  Minobe, Y.
  Nucleotide sequence of sequence tagged site from japonica rice
  Nipponbare as an RFLP marker
  Unpublished (1993)
  Submitted (18-Nov-1993) to DDBJ by:
  Yuzo Minobe
  Dept. Rice Genome Research Program
  National Institute of Agrobiological Resources
  Kannondai 2-1-2
  Tsukuba, Ibaraki
  Japan
  Phone: 0298-38-7441
  Fax: 0298-38-7468
  PROJECT "RGP".
FEATURES
  source
    Location/Qualifiers
      1..288
        /organism="Oryza sativa"
        /strain="Nipponbare"
        /db_xref="taxon:4530"
BASE COUNT      77 a 61 c 80 g 68 t 2 others
ORIGIN
Query Match      100.0%; Score 21; DB 11; Length 288;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21
    |||||||
Db 113 GTAACATAGATGACACCGCGC 93

RESULT 12
R1C1429B/c
LOCUS
DEFINITION Rice genomic DNA, L429B, sequence tagged site.
ACCESSION D25449
VERSION D25449.1 GI:436695
KEYWORDS STS; Not I linking clone; RFLP; Rice.
SOURCE Oryza sativa (strain:Nipponbare) DNA.
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
  1 (bases 1 to 288)
  Minobe, Y.
  Direct Submission
  Submitted (18-NOV-1993) to the DDBJ/EMBL/GenBank databases. Yuzo
  Minobe, National Institute of Agrobiological Resources, Rice Genome
  Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305, Japan
  (E-mail:MINOBE@rsg.riken.go.jp, Tel:0298-38-7441,
  Fax:0298-38-7468)
REFERENCE
  2 (bases 1 to 288)
  Minobe, Y.
  Nucleotide sequence of sequence tagged site from japonica rice
  Nipponbare as an RFLP marker
  Unpublished (1993)
  Submitted (18-Nov-1993) to DDBJ by:
  Yuzo Minobe
  Dept. Rice Genome Research Program
  National Institute of Agrobiological Resources
  Kannondai 2-1-2
  Tsukuba, Ibaraki
  Japan
  Phone: 0298-38-7441
  Fax: 0298-38-7468
  PROJECT "RGP".
FEATURES
  source
    Location/Qualifiers
      1..288
        /organism="Oryza sativa"
        /strain="Nipponbare"
        /db_xref="taxon:4530"
BASE COUNT      77 a 61 c 80 g 68 t 2 others
ORIGIN
Query Match      100.0%; Score 21; DB 11; Length 288;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21
    |||||||
Db 113 GTAACATAGATGACACCGCGC 93

RESULT 12
R1C1429B/c
LOCUS
DEFINITION Rice genomic DNA, L429B, sequence tagged site.
ACCESSION D25449
VERSION D25449.1 GI:436695
KEYWORDS STS; Not I linking clone; RFLP; Rice.
SOURCE Oryza sativa (strain:Nipponbare) DNA.
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
  1 (bases 1 to 288)
  Minobe, Y.
  Direct Submission
  Submitted (18-NOV-1993) to the DDBJ/EMBL/GenBank databases. Yuzo
  Minobe, National Institute of Agrobiological Resources, Rice Genome
  Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305, Japan
  (E-mail:MINOBE@rsg.riken.go.jp, Tel:0298-38-7441,
  Fax:0298-38-7468)
REFERENCE
  2 (bases 1 to 288)
  Minobe, Y.
  Nucleotide sequence of sequence tagged site from japonica rice
  Nipponbare as an RFLP marker
  Unpublished (1993)
  Submitted (18-Nov-1993) to DDBJ by:
  Yuzo Minobe
  Dept. Rice Genome Research Program
  National Institute of Agrobiological Resources
  Kannondai 2-1-2
  Tsukuba, Ibaraki
  Japan
  Phone: 0298-38-7441
  Fax: 0298-38-7468
  PROJECT "RGP".
FEATURES
  source
    Location/Qualifiers
      1..288
        /organism="Oryza sativa"
        /strain="Nipponbare"
        /db_xref="taxon:4530"
BASE COUNT      77 a 61 c 80 g 68 t 2 others
ORIGIN

```

```

ABA7623
LOCUS
DEFINITION Agrobacterium tumefaciens nos terminator with artificial insert of
            40 bp.
ACCESSION AJ007623
VERSION AJ007623.1 GI:3319860
KEYWORDS nos gene; terminator.
SOURCE Agrobacterium tumefaciens.
ORGANISM
  Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
  Rhizobiaceae; Rhizobium.
  1 (bases 1 to 318)
  Hardegger, M.R., Brodmann, P. and Herrmann, A.
  Quantitative detection of the 35S promoter and the NOS terminator
  using quantitative competitive PCR
  Z. Lebensm.-Unters. -Forsch., A Eur. Food Res. Technol. 209, 83-87
  (1999)
REFERENCE
  2 (bases 1 to 318)
  Hardegger, M.R.
  Direct Submission
  Submitted (03-JUL-1998) Hardegger M.R., Lebensmittel, Kantonales
  Laboratorium Basel-Stadt, Kannenfeldstrasse 2, CH/4012 Basel,
  SWITZERLAND
FEATURES
  source
    Location/Qualifiers
      1..318
        /organism="Agrobacterium tumefaciens"
        /plasmid="Ti"
        /db_xref="taxon:358"
      1..318
        /gene="nos"
      1..318
        /gene="nos"
      /note="Insert location 142-181"
BASE COUNT      102 a 64 c 55 g 97 t
ORIGIN
Query Match      100.0%; Score 21; DB 1; Length 318;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacccgcg 21
    |||||||
Db 16 GTAACATAGATGACACCGCGC 36

RESULT 13
ABA7624
LOCUS
DEFINITION Agrobacterium tumefaciens nos terminator with artificial insert of
            80 bp.
ACCESSION AJ007624
VERSION AJ007624.1 GI:3319861
KEYWORDS nos gene; terminator.
SOURCE Agrobacterium tumefaciens.
ORGANISM
  Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
  Rhizobiaceae; Rhizobium.
  1 (bases 1 to 358)
  Hardegger, M.R., Brodmann, P. and Herrmann, A.
  Quantitative detection of the 35S promoter and the NOS terminator
  using quantitative competitive PCR
  Z. Lebensm.-Unters. -Forsch., A Eur. Food Res. Technol. 209, 83-87
  (1999)
REFERENCE
  2 (bases 1 to 358)
  Hardegger, M.R.
  Direct Submission
  Submitted (03-JUL-1998) Hardegger M.R., Lebensmittel, Kantonales
  Laboratorium Basel-Stadt, Kannenfeldstrasse 2, CH/4012 Basel,
  SWITZERLAND
FEATURES
  source
    Location/Qualifiers
      1..358
        /organism="Agrobacterium tumefaciens"

```

```

/plasmid="Ti"
/db_xref="taxon:358"
1. .358
/gene="nos"
terminator
1. .358
/gene="nos"
/notes="insert location 142-221; direct repeat"
114 a 75 c 106 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 21; DB 1; Length 358;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacacgcgc 21
|||||
Db 16 GTAACATAGATGACACGC GC 36

```

RESULT	14
A71436/c	
LOCUS	A71436 563 bp DNA
DEFINITION	Sequence 6 from Patent WO9810081.
ACCESSION	A71436
VERSION	A71436.1 GI:4775049
KEYWORDS	.
SOURCE	unidentified.
ORGANISM	unidentified
REFERENCE	unclassified.
AUTHORS	1 (bases 1 to 563)
TITLE	Michiels,F. and Williams,M.
JOURNAL	IMPROVED BARSTAR GENE
FEATURES	Patent: WO 9810081-A 6 12-MAR-1998;
	MICHIELS FRANK (BE)
	Location/Qualifiers
source	1..563
	/organism="unidentified"
	/db_xref="taxon:32644"
BASE COUNT	152 a 133 c 145 g 133 t
ORIGIN	

```

Query Match      100.0%; Score 21; DB 6; Length 563;
Best Local Similarity 100.0%; Pred. NO. 0.25;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      gtaacatagatgacaccgcgc 21
          ||||||||||
Db       538    GTAACATAGATGACACCGCGC 518

RESULT 15
LOCUS   AX014764/c      838 bp      DNA      PAT      07-SEP-2000
DEFINITION
Sequence 19 from Patent WO9953053.
ACCESSION
AX014764
VERSION
AX014764.1 GI:10041035
KEYWORDS
.
SOURCE  synthetic construct.
        synthetic construct
        artificial sequence.
        1 (bases 1 to 838)
REFERENCE
AUTHORS  Lamberty,M., Hofmann,J., Bulet,P. and Brookhart,G.L.
TITLE     Gene coding for heliomicine and use thereof
          Patent: WO 9953053-A 19 21-OCT-1999;
          LAMBERTY MIREILLE (FR); HOFMANN JULES (FR);
          RHONE POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES
        source      1..838
                   /organism="synthetic construct"
                   /db_xref="taxon:32630"
        promoter    7..532

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 13:39:22 ; Search time 716.55 Seconds  
(without alignments)  
25.126 Million cell updates/sec

Title: US-09-698-903B-2  
Perfect score: 21  
Sequence: 1 gtaacatagatgacacggc 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	21	100.0	21	AAD06991	PCR primer MDB355
2	21	100.0	249	AAAF89820	Nucleotide sequenc
3	21	100.0	249	AAD06313	Nopaline synthetas
4	21	100.0	249	AAD06367	3' nopaline synth
5	21	100.0	249	AAAF89824	Nucleotide sequenc
6	21	100.0	249	AAAF89833	Nucleotide sequenc
7	21	100.0	253	AAZ51175	3'UTR of nopaline
8	21	100.0	254	AAZ98689	Nos terminator nuc
9	21	100.0	257	AAZ51653	Agrobacterium nos
10	21	100.0	261	AAAF5461	Nucleotide sequenc
11	21	100.0	261	AAD06309	Nopaline synthetas

c 12	21	100.0	261	AAD06363	3' nopaline syntha
c 13	21	100.0	261	AAAF89820	Nucleotide sequenc
c 14	21	100.0	261	AAAF89829	Nucleotide sequenc
c 15	21	100.0	263	AAAF89829	Nopaline synthase
c 16	21	100.0	270	AAAF89829	NOS terminator seq
c 17	21	100.0	270	AAAF89829	Tomato spotted wil
c 18	21	100.0	563	AAAF89829	Part of plasmid pL
c 19	21	100.0	900	AAAF89829	vector plasmid pSI
c 20	21	100.0	1138	AAAF89829	PMJBI DNA sequence
c 21	21	100.0	1138	AAAF89829	PMJBI DNA sequence
c 22	21	100.0	1138	AAAF89829	Plasmid PMJBI used
c 23	21	100.0	1138	AAAF89829	Plasmid PMJBI used
c 24	21	100.0	1559	AAAF89829	Arabidopsis thalia
c 25	21	100.0	1630	AAAF89829	Arabidopsis thalia
c 26	21	100.0	1690	AAAF89829	Expression cassett
c 27	21	100.0	1829	AAAF89829	Nematode-responsiv
c 28	21	100.0	1863	AAAF89829	35S Camv promoter-
c 29	21	100.0	1863	AAAF89829	Sequence of chimer
c 30	21	100.0	1923	AAAF89829	Tomato-tobacco end
c 31	21	100.0	1949	AAAF89829	OBPILT sequence.
c 32	21	100.0	2115	AAAF89829	Nucleotide sequenc
c 33	21	100.0	2145	AAAF89829	Oleosin-hirudin fu
c 34	21	100.0	2275	AAAF89829	Raspberry drul gen
c 35	21	100.0	2345	AAAF89829	Oligonucleotide #1
c 36	21	100.0	2345	AAAF89829	Arabidopsis thalia
c 37	21	100.0	2345	AAAF89829	Nucleotide sequenc
c 38	21	100.0	2378	AAAF89829	T-DNA sequence of
c 39	21	100.0	2436	AAAF89829	Agrobacterium tume
c 40	21	100.0	2543	AAAF89829	Expression cassett
c 41	21	100.0	2709	AAAF89829	Expression cassett
c 42	21	100.0	2917	AAAF89829	Oleosin-protein A
c 43	21	100.0	3069	AAAF89829	Raspberry drul gen
c 44	21	100.0	3121	AAAF89829	pND3018 repression
c 45	21	100.0	3183	AAAF89829	pND3008 activation

ALIGNMENTS

RESULT 1	
ID AAD06991	standard; DNA; 21 BP.
AC AAD06991;	
DT 06-AUG-2001	(first entry)
DE PCR primer MDB355 to amplify a fragment of pTCO113.	
XX	
XX Plasmid pTCO113; transgenic Brassica plant; transformation event;	
XX male-sterility gene; tobacco; PCR primer; probe; ss.	
XX Chimeric - Bacillus amyloliquefaciens.	
XX Chimeric - Nicotiana tabacum.	
XX WO200131042-A2.	
XX 03-MAY-2001.	
XX 26-OCT-2000; 2000WO-EP10680.	
XX 29-OCT-1999; 99US-0430497.	
XX (AVET ) AVENTIS CROPS SCIENCE NV.	
XX Weston B, De Beuckeleer M;	
XX WPI; 2001-300517/31.	
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by	
XX harboring specific transformation events, particularly by presence of	
XX male-sterility gene, at specific location in its genome -	

PS Claim 1; Page 26; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its seed, cells or tissues, characterised by harbouring a specific transformation event, particularly by the presence of a male-sterility gene, at a specific location in the Brassica genome. Transgenic Brassica plant is useful for producing a hybrid seed by crossing the transgenic plant with a male-fertile Brassica plant and harvesting the hybrid seed from the transgenic Brassica plant.

CC The present sequence is PCR primer WDB355 used to amplify a fragment of plasmid pCOL13. The primer is also used as a probe. The amplified fragment comprises coding region of barnase from *Bacillus amyloliquefaciens* and promoter region of the anther-specific gene TA29 from *Nicotiana tabacum*. This primer corresponds to position 2667-2687 of plasmid pCOL13.

XX Sequence 21 BP; 7 A; 6 C; 5 G; 3 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
|||||  
DB 1 gtaacatagatgacacgcgc 21

RESULT 2  
AA75465/C  
ID AA75465 standard; DNA; 249 BP.  
XX AA75465;  
AC  
DT 15-JAN-2001 (first entry)  
DE Nucleotide sequence of a nopaline synthase terminator.  
XX  
KW 35S promoter; alcohol dehydrogenase intron; nopaline synthase terminator;  
KW insecticidal Cry IAb protein toxin; transgenic plant; fungal toxin;  
KW aflatoxin; baculovirus; Lepidopteran insect; glufosinate; ds.  
XX Unidentified.  
XX US6114608-A.  
PN  
XX 05-SEP-2000.  
XX 13-MAR-1998; 9805-0042426.  
XX 14-MAR-1997; 9705-0109808.  
XX (NOVS ) NOVARTIS AG.  
PA Dietrich PS, Mettler IJ, Sinibaldi RM;  
PI WPI; 2000-586487/55.  
DR  
XX Novel DNA construct comprising an expression cassette having a functional constitutive promoter, operably linked to a maize alcohol dehydrogenase intron, gene encoding Cry IAb protein and a terminator -

Claim 2; Column 25-26; 24pp; English.

XX The present sequence represents a nopaline synthase terminator, which is used to create the construct of the invention. The specification describes a nucleic acid construct, comprising an expression cassette containing a cauliflower mosaic virus (CaMV) 35S promoter, a maize alcohol dehydrogenase intron, a DNA molecule encoding an insecticidal Cry IAb protein toxin and a NOS (nopaline synthase) terminator in operable linkage. The nucleic acid constructs are useful for producing transgenic plants such as maize, wheat, barley, sorghum and rice. Preferably, maize plants such as field corn, sweet corn, white

CC corn, silage corn or popcorn are produced. The transgenic maize is used for preparing food materials with reduced levels of fungal toxins, e.g. aflatoxins. The transgenic plants express a baculovirus crystal protein toxin, which is effective against lepidopteran insects at relatively high levels and further provides resistance to the non-selective herbicide glufosinate.

XX Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
|||||  
DB 247 GTAACATAGATGACACGC GC 227

RESULT 3  
AAD06313/C  
ID AAD06313 standard; DNA; 249 BP.  
XX AAD06313;  
AC  
DT 10-AUG-2001 (first entry)  
DE Nopaline synthase terminator of Pat gene expression cassette.  
XX  
KW Transgenic maize; expression cassette; 35S promoter; nopaline synthase;  
KW NOS terminator; alcohol dehydrogenase intron; CryIAb toxin;  
KW foodstuffs preparation; phosphinothricin acetyl transferase; PAT; ds.  
XX Unidentified.  
XX US6222104-B1.  
PN  
XX 24-APR-2001.  
XX 13-APR-1999; 99US-0291238.  
XX 09-NOV-1994; 94US-0336627.  
PR 22-AUG-1996; 96US-0716836.  
PR 14-MAR-1997; 97US-0818573.  
PR 13-MAR-1998; 98US-0042426.  
XX (NOVS ) NOVARTIS AG.  
PA Mettler IJ, Krier M, Miles D;  
PI WPI; 2001-327266/34.  
DR  
XX Novel transgenic maize seed for hybrid maize plant production, comprising expression cassette linked operably with CaMV 35S promoter, alcohol dehydrogenase intron, insecticidal CryIAb toxin gene or NOS terminator -

Claim 3; Column 25-26; 24pp; English.

XX The present invention relates to transgenic maize seed for hybrid maize plant production, comprising an expression cassette linked operably with constitutive Cauliflower mosaic virus (CaMV) 35S promoter, an alcohol dehydrogenase intron, a DNA encoding an insecticidal *Bacillus thuringiensis* (Bt) CryIAb toxin and a nopaline synthase (NOS) terminator. The transgenic maize seed is used in maize cultivation methods for propagating hybrid maize seeds and for growing hybrid maize plants. The maize is also used in foodstuffs preparation for animal or human consumption. The inbred maize line NP948 of the invention has reduced level of fungal toxins, hence suitable for foodstuffs preparation. Yield is high. Plant health is improved due to reduced stalk rot. Grain test weight is high and the rate of grain dry down is reduced. The present sequence is nopaline synthase terminator of phosphinothricin acetyl transferase (PAT) gene expression cassette.

XX SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 4  
AAD06367/c  
ID AAD06367 standard; DNA; 249 BP.  
XX  
AC AAD06367;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE 3' nopaline synthase terminator of Pat gene expression cassette.  
XX  
KW Maize: expression cassette; 35S promoter; alcohol dehydrogenase intron;  
KW crystal 1ab; CryIAb toxin; phosphinothricin acetyl transferase; PAT;  
KW nopaline synthase; NOS terminator; aflatoxin; food material;  
KW sweet corn; human food; Adhl-1S; ds.  
XX  
OS Unidentified.  
XX  
PN US6229075-B1.  
XX  
PD 08-MAY-2001.  
XX  
PF 11-JUN-1999; 99US-0330760.  
XX  
PR 14-MAR-1997; 97US-0109808.  
PR 13-MAR-1998; 98US-0042426.  
XX  
PA (NOVS ) NOVARTIS AG.  
XX  
PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;  
XX  
DR WPI; 2001-342708/36.  
XX  
PT New maize (Zea mays) inbred line R412H (ATCC 209675), useful for  
PT producing corn with reduced levels of toxins (e.g. the fungal  
PT aflatoxin) and for preparing food materials for human or animal  
PT consumption -  
XX  
PS Claim 3; Column 21-22; 30pp; English.  
XX  
CC The present invention relates to maize inbred line R412H which comprises  
CC a nucleic acid construct with two cassettes, which are transcribed in the  
CC same direction. The expression cassette comprises a Cauliflower mosaic  
CC virus (CaMV) 35S constitutive promoter operably linked to a maize alcohol  
CC dehydrogenase intron, a DNA sequence of a gene encoding a  
CC Bacillus thuringiensis (Bt) crystal 1Ab (CryIAb) toxin or  
CC phosphinothricin acetyl transferase (PAT) and a nopaline synthase (NOS)  
CC terminator functional in plants. The maize inbred line R412H is useful  
CC for producing corn with reduced levels of toxins (e.g. the fungal  
CC aflatoxin). This maize line is particularly useful for preparing food  
CC materials for human or animal consumption, e.g. sweet corn for  
CC packaging or fresh use as human food, or grain or silage made from field  
CC corn. The present sequence is 3' NOS terminator of Pat gene expression  
CC cassette, which is related to the invention.  
XX  
SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 5  
AAF89824/c  
ID AAF89824 standard; DNA; 249 BP.  
XX  
AC AAF89824;  
XX  
DT 23-JUL-2001 (first entry)  
XX  
DE Nucleotide sequence of a 3' terminator from nopaline synthetase.  
XX  
KW Maize; inbred line R660H; 35S constitutive promoter; CryIAb protein;  
KW alcohol dehydrogenase; phosphinothricin acetyl transferase; Chromosome 8;  
KW Z1B3; UMC150a; Lepidoptera; gluofosinate; sweet corn; fungal toxin; ds.  
XX  
OS Unidentified.  
XX  
PN US6232533-B1.  
XX  
PD 15-MAY-2001.  
XX  
PF 09-JUN-1999; 99US-0328473.  
XX  
PR 14-MAR-1997; 97US-0109808.  
PR 13-MAR-1998; 98US-0042426.  
XX  
PA (NOVS ) NOVARTIS AG.  
XX  
PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;  
XX  
DR WPI; 2001-335091/35.  
XX  
PT Novel seed of maize inbred line R372H, useful for producing inbred  
PT maize plants which on crossing with other different maize plants  
PT produce hybrid maize seeds and plants having resistance to insects and  
PT herbicide -  
XX  
PS Claim 3; Column 19-21; 29pp; English.  
XX  
CC The specification describes seed and plants of maize inbred line R660H.  
CC The seed comprises two cassettes. The first cassette comprises a  
CC cauliflower mosaic virus (CaMV) 35S constitutive promoter operably linked  
CC to a maize alcohol dehydrogenase intron, a DNA sequence of a gene  
CC encoding a CryIAb protein, and a terminator which functions in plants. The  
CC second cassette comprises a CaMV 35S promoter which functions in plant  
CC cells operably linked to a maize alcohol dehydrogenase intron, a DNA  
CC sequence of a gene encoding for phosphinothricin acetyl transferase, and  
CC a terminator functional in plant. The two cassettes are transcribed into  
CC the same direction and the nucleic acid construct is incorporated into  
CC the seed's genome on chromosome 8 and near position 117, between markers  
CC Z1B3 and UMC150a. The maize plants and seeds exhibit resistance to  
CC Lepidopteran insects since they express the protein CryIAb in high  
CC levels, and also exhibit resistance to non-selective herbicide  
CC glufosinate. The transgenic maize is suited for preparation of food  
CC materials for human or animal consumption e.g. sweet corn for packaging  
CC or fresh use as a human food, or grain or silage made from field corn,  
CC containing reduced levels of fungal toxins. The present sequence  
CC represents a 3' terminator from nopaline synthetase, which is used to  
CC produce transgenic maize of the invention.  
XX  
SQ Sequence 249 BP; 79 A; 38 C; 47 G; 85 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 gtaacatagatgacaccgcgc 21







```

PN US6114608-A.
XX
PD 05-SEP-2000.
XX
PF 13-MAR-1998; 98US-0042426.
XX
PR 14-MAR-1997; 97US-0109808.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Dietrich PS, Mettler IJ, Sinibaldi RM;
XX
DR WPI; 2000-586487/55.
XX
PT Novel DNA construct comprising an expression cassette having a
PT functional constitutive promoter, operably linked to a maize alcohol
PT dehydrogenase intron, gene encoding Cry IAb protein and a terminator
XX
PS Claim 1; Column 21-22; 24pp; English.
XX
CC The present sequence represents a nopaline synthase terminator, which is
CC used to create the construct of the invention. The specification
CC describes a nucleic acid construct, comprising an expression cassette
CC containing a cauliflower mosaic virus (CaMV) 35S promoter, a maize
CC alcohol dehydrogenase intron, a DNA molecule encoding an insecticidal
CC Cry IAb protein toxin and a NOS (nopaline synthase) terminator in
CC operable linkage. The nucleic acid constructs are useful for
CC producing transgenic plants such as maize, wheat, barley, sorghum and
CC rice. Preferably, maize plants such as field corn, sweet corn, white
CC corn, silage corn or popcorn are produced. The transgenic maize is
CC used for preparing food materials with reduced levels of fungal toxins,
CC e.g. aflatoxins. The transgenic plants express a baculovirus crystal
CC protein toxin, which is effective against Lepidopteran insects at
CC relatively high levels and further provides resistance to the
CC non-selective herbicide glufosinate.
XX
SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21
DB 253 GTAACATAGATGACACCGCGC 233

RESULT 11
AAD06309/c
XX AAD06309 standard; DNA; 261 BP.
XX
AC AAD06309;
XX
DT 10-AUG-2001 (first entry)
XX
DE Nopaline synthase terminator of Bt kurstaki expression cassette.
XX
KW Transgenic maize; expression cassette; 35S promoter;
KW alcohol dehydrogenase intron; CryIAb toxin; nopaline synthetase;
KW NOS terminator; foodstuffs preparation; ds.
XX
OS Unidentified.
XX
PN US6222104-B1.
XX
PD 24-APR-2001.
XX
PF 13-APR-1999; 99US-0291238.
XX
PR 09-NOV-1994; 94US-0336627.
PR 22-AUG-1996; 96US-0716836.
PR 14-MAR-1997; 97US-0818573.

```

```

PR 13-MAR-1998; 98US-0042426.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Mettler IJ, Krier M, Mies D;
XX
DR WPI; 2001-327266/34.
XX
PT Novel transgenic maize seed for hybrid maize plant production,
PT comprising expression cassette linked operably with CaMV 35S promoter,
PT alcohol dehydrogenase intron, insecticidal CryIAb toxin gene or NOS
PT terminator
XX
PS Claim 1; Column 21-22; 24pp; English.
XX
CC The present invention relates to transgenic maize seed for hybrid maize
CC plant production, comprising an expression cassette linked operably with
CC constitutive cauliflower mosaic virus (CaMV) 35S promoter, an alcohol
CC dehydrogenase intron, a DNA encoding an insecticidal
CC bacillus thuringiensis (Bt) CryIAb toxin and a nopaline synthetase (NOS)
CC terminator. The transgenic maize seed is used in maize cultivation
CC methods for propagating hybrid maize seeds and for growing hybrid maize
CC plants. The maize is also used in foodstuffs preparation for animal or
CC human consumption. The inbred maize line NP948 of the invention has
CC reduced level of fungal toxins, hence suitable for foodstuffs
CC preparation. Yield is high. Plant health is improved due to reduced
CC stalk rot. Grain test weight is high and the rate of grain dry down is
CC reduced. The present sequence is nopaline synthetase terminator of Bt
CC kurstaki expression cassette.
XX
SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 261;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21
DB 253 GTAACATAGATGACACCGCGC 233

RESULT 12
AAD06363/c
XX AAD06363 standard; DNA; 261 BP.
XX
AC AAD06363;
XX
DT 10-AUG-2001 (first entry)
XX
DE 3' nopaline synthase terminator of Btk gene expression cassette.
XX
KW Maize; expression cassette; 35S promoter; alcohol dehydrogenase intron;
KW crystal IAb; CryIAb toxin; phosphinothricin acetyl transferase; PAT;
KW nopaline synthase; NOS terminator; aflatoxin; food material;
KW sweet corn; human food; Adhl-1S; ds.
XX
OS Unidentified.
XX
PN US6229075-B1.
XX
PD 08-MAY-2001.
XX
PF 11-JUN-1999; 99US-0330760.
XX
PR 14-MAR-1997; 97US-0109808.
PR 13-MAR-1998; 98US-0042426.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Mettler IJ, Plaisted DC, Grier SL, Houghton W, Gardiner M;
XX
DR WPI; 2001-342708/36.

```



CC cauliflower mosaic virus (CaMV) 35S constitutive promoter operably linked  
 CC to a maize alcohol dehydrogenase intron, a DNA sequence of a gene  
 CC encoding a CryIAb protein, and a terminator functional in plants. The  
 CC second cassette comprises a CaMV 35S promoter which functions in plant  
 CC cells operably linked to a maize alcohol dehydrogenase intron, a DNA  
 CC sequence of a gene encoding for phosphinothricin acetyl transferase, and  
 CC a terminator functional in plant. The two cassettes are transcribed into  
 CC the same direction and the nucleic acid construct is incorporated into  
 CC the seed's genome on chromosome 8 and near position 117, between markers  
 CC Z1B3 and UMC150a. The maize plants and seeds exhibit resistance to  
 CC Lepidopteran insects since they express the protein CryIAb in high  
 CC levels and also exhibit resistance to non-selective herbicide  
 CC glufosinate. The transgenic maize is suited for preparation of food  
 CC materials for human or animal consumption e.g. sweet corn for packaging  
 CC or fresh use as a human food, or grain or silage made from field corn,  
 CC containing reduced levels of fungal toxins. The present sequence  
 CC represents a 3' terminator from nopaline synthetase, which is used to  
 CC produce transgenic maize of the invention.

SQ Sequence 261 BP; 83 A; 41 C; 50 G; 87 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 |||||  
 DB 253 GTAACATAGATGACACGCGC 233

RESULT 15  
 AAT66536/c  
 ID AAT66536 standard; DNA; 263 BP.

AC AAT66536;

XX 22-JUL-1997 (first entry)

DT Nopaline synthase 3'nontranslated region.

DE Transgenic plant; selectable marker; carotenoid; pigment;  
 KW nopaline synthase; phytoene synthase; plasmid pET0203; ss.

XX Agrobacterium sp.

OS WO9714807-A1.

PN 24-APR-1997.

PD 29-MAR-1996; 96WO-US04313.

PF 16-OCT-1995; 95US-0543608.

XX (SEMI-) SEMINIS VEGATABILES.

XX Braun CJ, Trulson AJ;

PI WPI; 1997-245122/22.

XX Visual identification of transgenic plant material - from production  
 XX of carotenoid pigment encoded by cassette containing Erwinia  
 XX phytoene synthase gene, useful for selecting material for  
 XX regeneration

PS Example; Page 37; 62pp; English.

XX The 3' nontranslated region (AAT66536) of the nopaline synthase  
 CC gene can be cloned from binary vector pB1121 (Clontech). It  
 CC contains sequences that in plant cells result in the termination  
 CC of transcription and additional sequences that when transcribed  
 CC into RNA result in the addition of a poly-A tract to the 3' end  
 CC of the RNA. It is a preferred component of pET0203 (ATCC 97282), a

CC binary vector used in a method for the visual identification of  
 CC transgenic plant material. The vector includes an expression  
 CC cassette comprising the tomato P8 promoter (see also AAT66533), a  
 CC plastid targeting signal (AAT66535) fused to the Erwinia herbicola  
 CC phytoene synthase gene (AAT66534), and the 3' non-translated region.  
 CC Transgenic plant cells and tissues are identified by the  
 XX appearance of orange colour.

SQ Sequence 263 BP; 83 A; 42 C; 49 G; 89 T; 0 other;

Query Match 100.0%; Score 21; DB 18; Length 263;  
 Best Local Similarity 100.0%; Pred. No. 0.086;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21  
 |||||  
 DB 257 GTAACATAGATGACACGCGC 237

Search completed: February 25, 2002, 18:17:19  
 Job time: 16677 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:55:52 ; Search time 301.6 Seconds  
(without alignments)  
15.769 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21  
Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	21	100.0	249	3	US-09-042-426-8
C 2	21	100.0	249	4	US-09-291-238-8
C 3	21	100.0	249	4	US-09-330-760-8
C 4	21	100.0	249	4	US-09-328-473-8
C 5	21	100.0	249	4	US-09-330-737-8
C 6	21	100.0	254	3	US-08-869-696-18
C 7	21	100.0	261	3	US-09-042-426-4
C 8	21	100.0	261	4	US-09-291-238-4
C 9	21	100.0	261	4	US-09-330-760-4
C 10	21	100.0	261	4	US-09-328-473-4
C 11	21	100.0	261	4	US-09-330-737-4
C 12	21	100.0	1287	1	US-08-064-121-3
C 13	21	100.0	1287	1	US-08-478-015-3
C 14	21	100.0	1287	3	US-08-475-975-3
C 15	21	100.0	1287	3	US-09-084-889-3
C 16	21	100.0	1559	3	US-09-049-475-6
C 17	21	100.0	1829	1	US-07-966-187-17
C 18	21	100.0	1863	1	US-08-525-507-16
C 19	21	100.0	1863	2	US-08-475-427-9
C 20	21	100.0	1863	2	US-07-842-165-9
C 21	21	100.0	2115	2	US-08-767-026-3
C 22	21	100.0	2145	1	US-08-592-936B-16
C 23	21	100.0	2145	2	US-09-111-573-16
C 24	21	100.0	2319	3	US-09-049-475-5
C 25	21	100.0	2345	3	US-09-026-673-1
C 26	21	100.0	2633	1	US-08-452-267-2
C 27	21	100.0	2633	3	US-09-123-644-2

Sequence 20, Appl  
Sequence 20, Appl  
Sequence 1, Appl  
Sequence 26, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 14, Appl  
Sequence 9, Appl  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl

US-09-042-426-8/c  
; Sequence 8, Application US/09042426  
; Patent No. 6114608  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6114608artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
; US-09-042-426-8

#### ALIGNMENTS

RESULT 1  
US-09-042-426-8/c  
; Sequence 8, Application US/09042426  
; Patent No. 6114608  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6114608artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
; US-09-042-426-8

Query Match 100.0%; Score 21; DB 3; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21

```
Db 247 GTAACATAGATGACACCGCGC 227
|||||
RESULT 2
US-09-291-238-8/c
; Sequence 8, Application US/09291238
; Patent No. 6222104
; GENERAL INFORMATION:
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6222104artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/291,238
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS Terminator
; US-09-291-238-8

Query Match 100.0%; Score 21; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 3
US-09-330-760-8/c
; Sequence 8, Application US/09330760
; Patent No. 6229075
; GENERAL INFORMATION:
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6229075artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/328,473
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS Terminator
; US-09-330-760-8

Query Match 100.0%; Score 21; DB 4; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 247 GTAACATAGATGACACCGCGC 227

RESULT 4
US-09-328-473-8/c
; Sequence 8, Application US/09328473
; Patent No. 6232533
; GENERAL INFORMATION:
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
; TITLE OF INVENTION: DNA Construct Containing Bacillus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6232533artis Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/328,473
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,426
; FILING DATE: March 13, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoxie, Thomas
; REGISTRATION NUMBER: 32,993
; REFERENCE/DOCKET NUMBER: 135/1
; TELEPHONE: (919) 541-8614
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: NOS Terminator
; US-09-330-760-8
```

; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: NOS Terminator  
; US-09-328-473-8

Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 247 GTAACATAGATGACACCGGC 227

RESULT 5  
US-09-330-737-8/c  
; Sequence 8, Application US/09330737  
; Patent No. 6232534  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6232534artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/09/330,737  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993  
; REFERENCE/DOCKET NUMBER: 135/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 541-8614  
; TELEFAX: (919) 541-8689  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:

; CLONE: NOS Terminator  
; US-09-330-737-8

Query Match 100.0%; Score 21; DB 4; Length 249;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 247 GTAACATAGATGACACCGGC 227

RESULT 6  
US-08-869-696-18/c  
; Sequence 18, Application US/08869696C  
; Patent No. 6031155  
; GENERAL INFORMATION:  
; APPLICANT: Cameron-Mills, Verena  
; APPLICANT: Lok, Finn  
; APPLICANT: Sinjorgo, Catharina Maria Cornelia  
; APPLICANT: Van Den Dool, Ronald Tako Marinus  
; APPLICANT: Caspers, Martinus Petrus Maria  
; APPLICANT: Van Zuijl-Van Der Valk, Maria Joanna  
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION  
; FILE REFERENCE: 11225.01US01  
; CURRENT APPLICATION NUMBER: US/08/869,696C  
; CURRENT FILING DATE: 1997-06-05  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 254  
; TYPE: DNA  
; ORGANISM: synthetic  
; US-08-869-696-18

Query Match 100.0%; Score 21; DB 3; Length 254;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 247 GTAACATAGATGACACCGGC 227

RESULT 7  
US-09-042-426-4/c  
; Sequence 4, Application US/09042426  
; Patent No. 6114608  
; GENERAL INFORMATION:  
; APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
; TITLE OF INVENTION: DNA Construct Containing Bacillus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6114608artis Corporation  
; STREET: 564 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/09/042,426  
; FILING DATE: March 13, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoxie, Thomas  
; REGISTRATION NUMBER: 32,993

REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-042-426-4

Query Match 100.0%; Score 21; DB 3; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 8  
US-09-291-238-4/c  
Sequence 4, Application US/09291238  
Patent No. 622104  
GENERAL INFORMATION:  
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
TITLE OF INVENTION: DNA Construct Containing Bacillus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 622104artis Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/291,238  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/042,426  
FILING DATE: March 13, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoxie, Thomas  
REGISTRATION NUMBER: 32,993  
REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-291-238-4

Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 9  
US-09-330-760-4/c  
Sequence 4, Application US/09330760  
Patent No. 6229075  
GENERAL INFORMATION:  
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
TITLE OF INVENTION: DNA Construct Containing Bacillus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6229075artis Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/330,760  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/042,426  
FILING DATE: March 13, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoxie, Thomas  
REGISTRATION NUMBER: 32,993  
REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-330-760-4  
Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 GTAACATAGATGACACCGCGC 233

RESULT 10  
US-09-328-473-4/c  
Sequence 4, Application US/09328473  
Patent No. 6232533  
GENERAL INFORMATION:



APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
TITLE OF INVENTION: DNA Construct Containing Bacillus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 623253artis Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/328,473  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 09/042,426  
FILING DATE: March 13, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoxie, Thomas  
REGISTRATION NUMBER: 32,993  
REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-328-473-4

Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 gtaacatagatgacaccgcgc 233

RESULT 11  
US-09-330-737-4/c  
Sequence 4, Application US/09330737  
Patent No. 6232534  
GENERAL INFORMATION:  
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi  
TITLE OF INVENTION: DNA Construct Containing Bacillus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6232534artis Corporation  
STREET: 564 Morris Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/330,737  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/09/042,426  
FILING DATE: March 13, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoxie, Thomas  
REGISTRATION NUMBER: 32,993  
REFERENCE/DOCKET NUMBER: 135/1  
TELEPHONE: (919) 541-8614  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
IMMEDIATE SOURCE:  
CLONE: NOS terminator  
US-09-330-737-4

Query Match 100.0%; Score 21; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 0.047;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21  
|||||  
Db 253 gtaacatagatgacaccgcgc 233

RESULT 12  
US-08-064-121-3/c  
Sequence 3, Application US/08064121  
Patent No. 5641664  
GENERAL INFORMATION:  
APPLICANT: D'HALLUIN, Kathleen  
APPLICANT: Gobel, Elke  
TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/064,121  
APPLICATION NUMBER: US/08/064,121  
FILING DATE: 24-MAY-1993  
CLASSIFICATION: 800  
Prior Application Data:  
APPLICATION NUMBER: EP 90403332.1  
FILING DATE: 23-NOV-1990  
Prior Application Data:  
APPLICATION NUMBER: EP 91401888.2  
FILING DATE: 08-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 010830-043  
TELECOMMUNICATION INFORMATION:

```
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
; US-08-064-121-3

Query Match 100.0%; Score 21; DB 1; Length 1287;
Best Local Similarity 100.0%; Pred. NO. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 1251 GTAACATAGATGACACGC GC 1231
```

```
RESULT 13
US-08-478-015-3/c
; Sequence 3, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylat
; OTHER INFORMATION: site derived from Agrobacterium T-DNA nopaline synthase ge
; US-08-478-015-3

Query Match 100.0%; Score 21; DB 1; Length 1287;
Best Local Similarity 100.0%; Pred. NO. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21
|||||
Db 1251 GTAACATAGATGACACGC GC 1231

RESULT 14
US-08-475-975-3/c
; Sequence 3, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475.975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 9040332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= pTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
;
US-08-475-975-3

Query Match 100.0%; Score 21; DB 3; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
|||||
Db 1251 GTAACATAGATGACACCGCGC 1231

RESULT 15
US-09-084-889-3/c
; Sequence 3, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.

```

```

;
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,121
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: fragment of pTTM8
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..545
; OTHER INFORMATION: /label= pTA29
; OTHER INFORMATION: /note= "Promoter from the TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 546..881
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding sequence of barnase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 882..1287
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene "
;
US-09-084-889-3

Query Match 100.0%; Score 21; DB 3; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaacatagatgacaccgcgc 21
|||||
Db 1251 GTAACATAGATGACACCGCGC 1231

Search completed: February 25, 2002, 18:05:08
Job time: 18556 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:51:47 ; Search time 8261.74 Seconds  
(without alignments)  
27.314 Million cell updates/sec

Title: US-09-698-903B-2

Perfect score: 21

Sequence: 1 gtaacatagatgacaccgcgc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estfun:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estom:\*
  - 5: em\_estpl:\*
  - 6: em\_estba:\*
  - 7: em\_estro:\*
  - 8: em\_estov:\*
  - 9: em\_htc:\*
  - 10: gb\_est1:\*
  - 11: gb\_est2:\*
  - 12: gb\_htc:\*
  - 13: gb\_gss:\*
  - 14: em\_gss\_fun:\*
  - 15: em\_gss\_hum:\*
  - 16: em\_gss\_inv:\*
  - 17: em\_gss\_pln:\*
  - 18: em\_gss\_pro:\*
  - 19: em\_gss\_rod:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
c	1	17.8	84.8	887	13	AQ363458
c	2	17.4	82.9	300	10	AU114795
c	3	17.4	82.9	833	11	BF104658
c	4	17.4	82.9	936	13	CNS01EY4
c	5	17	81.0	215	10	BB143973
c	6	17	81.0	277	10	BB244247
c	7	16.8	80.0	232	10	AV008164
c	8	16.8	80.0	379	10	AV684098
c	9	16.8	80.0	542	13	CNS00WTL
c	10	16.8	80.0	594	10	AI109780
c	11	16.8	80.0	662	13	AZ029768
c	12	16.8	80.0	710	10	BE038757

13																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

ALIGNMENTS

RESULT 1  
LOCUS AQ363458/c  
DEFINITION nbxb0059A08f CUGI Rice BAC Library Oryza sativa genomic clone  
ACCESSION AQ363458  
VERSION AQ363458.2  
KEYWORDS GSS, GI:6583332  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.  
REFERENCE 1 (bases 1 to 887)  
AUTHORS Wing, R.A. and Dean, R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT On Dec 15, 1999 this sequence version replaced gi:4213113.  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACATAGGG  
Class: BAC ends  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .887  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbxb0059A08f"

AQ363458 887 bp DNA GSS 16-DEC-1999  
nbxb0059A08f CUGI Rice BAC Library Oryza sativa genomic clone  
nbxb0059A08f, DNA sequence.  
AQ363458  
AQ363458.2 GI:6583332  
GSS.  
Oryza sativa.  
Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 887)  
Wing, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
On Dec 15, 1999 this sequence version replaced gi:4213113.  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACATAGGG  
Class: BAC ends  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .887  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbxb0059A08f"

/clone\_lib="CUGI Rice BAC Library"

/tissue\_type="Leaf"

/lab\_host="E. coli DH10B"

/notes="Vector: pBelBAC11; Site\_1: HindIII; Site\_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

the whole library for colony screening."

260 a 200 c 189 g 236 t 2 others

Query Match 84.8%; Score 17.8; DB 13; Length 887;  
Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgcgc 21

||||| ||||||| |||||||

Db 517 GCACATTTGATGACACGC GC 497

#### RESULT 2

AU114795/c

LOCUS

DEFINITION AU114795 300 bp mRNA EST 19-OCT-2000  
elegans cDNA clone yk728e10 3', mRNA sequence.

ACCESSION AU114795

VERSION AU114795.1

KEYWORDS GI:10928362

SOURCE EST.

ORGANISM Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 300)

Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

and Sugano, S.

A complementary view of the C. elegans genome

Unpublished (2000)

Contact: Yuji Kohara

Genome Biology Lab.

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

Location/Qualifiers

1. .300

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk728e10"

/clone\_lib="unpublished oligo-capped cDNA library"

/sex="Hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="varied"

96 a 49 c 66 g 86 t 3 others

BASE COUNT

ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 300;  
Best Local Similarity 94.7%; Pred. No. 1.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gtaacatagatgacacgc 19

||||| ||||||| |||||||

Db 148 GTAACATAGATGACACGC 130

#### RESULT 3

BF104658/c

LOCUS

DEFINITION BF104658 833 bp mRNA EST 19-OCT-2000

601647679R1 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4041811 3',

mRNA sequence.

ACCESSION BF104658

VERSION BF104658.1

KEYWORDS GI:10887184

SOURCE EST.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 833)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM861 row: n column: 20.

Location/Qualifiers

1. .833

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4041811"

/clone\_lib="NIH\_MGC\_61"

/tissue\_type="embryonal carcinoma"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: Testis; Vector: pDNR-LIB (Clontech); Site\_1:

SfiI (ggccatagggc); Site\_2: SfiI (ggccatagggc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-ATTCTAGAGCGGCGGACATG-dt(30)BN-3'

sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor

(where B = A, C, G and N = A, C, G, or T). Average

insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC

Library."

199 a 213 c 196 g 225 t

BASE COUNT

ORIGIN

Query Match 82.9%; Score 17.4; DB 11; Length 833;  
Best Local Similarity 94.7%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 taacatagatgacacgcg 20

||||| ||||||| |||||||

Db 600 TAACATAGATGACCGCG 582

#### RESULT 4

CNS01EY4

LOCUS

DEFINITION CNS01EY4 936 bp DNA GSS 30-MAY-2001

Anopheles gambiae GSS T7 end of clone 03E09 of Notredamei library

from strain PEST of Anopheles gambiae (African malaria mosquito),



## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 277)  
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kawada, I., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata  
Y., Shigemoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suganara, Y.,  
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya  
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,  
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino  
M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome.res@gsr.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source

1. .277  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="A73001C13"  
/clone\_lib="RIKEN full-length enriched, 7 days neonate  
cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="7 days neonate"  
/lab\_host="DH10B"  
/note="Site 1: Salt; Site 2: BamHI: cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction to Rot = 459.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATCTCGAGTAAATTAATTAATCCGCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(-) after bulk excision from Lambda  
FLC I."

BASE COUNT  
ORIGIN

83 a 49 c 55 g 90 t

## Query Match

Best Local Similarity 81.0%; Score 17; DB 10; Length 277;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY 2

2 taacatagatgacacgcg 18

## DB 65

65 TAACATAGATGACACGC 81

## RESULT 7

## AV008164

## LOCUS

## DEFINITION

AV008164 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA

## ACCESSION

AV008164

## VERSION

AV008164.1 GI:4785151

## KEYWORDS

EST.

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

1 (bases 1 to 232)

## AUTHORS

Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara  
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,  
Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,  
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara  
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,  
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
RIKEN Mouse ESTs  
Unpublished (1999)  
Contact: Chie Owa  
Genome Science Laboratory  
RIKEN  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098  
Email: genome.res@rtc.riken.go.jp  
Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

1. .232

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="1110008H08"

/clone\_lib="Mus musculus 18-day embryo C57BL/6J"

/sex="mixed"

/dev\_stage="18-day embryo"

84 a 32 c 45 g 71 t

## Query Match

Best Local Similarity 80.0%; Score 16.8; DB 10; Length 232;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## QY 1

1 gtaacatagatgacacgcg 20

## DB 102

102 GTAACATAGATGACACGC 121

## RESULT 8

## AV684098

## LOCUS

## DEFINITION

AV684098 379 bp mRNA

## ACCESSION

AV684098

EST  
25-SEP-2000

GKC Homo sapiens cDNA clone GKBCG09 5', mRNA sequence.



```

VERSION AV584098.1 GI:10285961
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarhini; Homnidae; Homo.
1 (bases 1 to 379)
Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GK- clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Source
1..379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBQ09"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOUR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 89 a 80 c 87 g 123 t
ORIGIN
Query Match 80.0%; Score 16.8; DB 10; Length 379;
Best Local Similarity 90.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 2 taacatagatgacaccgcgc 21
Db 352 TCACATAGATGACACCGCCC 371
RESULT 9
CNS00WTL 542 bp DNA GSS 28-JUN-1999
LOCUS Arabidopsis thaliana genome survey sequence T7 end of BAC T12G21 of
DEFINITION TAMU library from strain Columbia of Arabidopsis thaliana, genomic
survey sequence.
ACCESSION AL094023.1 GI:5295177
VERSION AL094023.1 GI:5295177
KEYWORDS GSS.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicaceae; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 542)
AUTHORS Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P.,
Samson,D., Saurin,W., Weissbach,J. and Quetier,F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 542)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
Source
1..542
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="TAMU"
/clone="T12G21"
/note="end : T7"
BASE COUNT 205 a 88 c 110 g 139 t
ORIGIN
Query Match 80.0%; Score 16.8; DB 13; Length 542;
Best Local Similarity 90.0%; Pred. No. 3.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 gtaacatagatgacaccgcgc 20
Db 518 GTAACATAGATTACACGGCG 537
RESULT 10
AL109780 594 bp mRNA EST 23-APR-2001
LOCUS GH09125.3prime GH Drosophila melanogaster head pOT2 Drosophila
DEFINITION melanogaster cDNA clone GH09125 5 similar to GS16904: FBan0016904
located on: 3R 85E13-85E13;; 04/10/2001, mRNA sequence.
ACCESSION AL109780
VERSION AL109780.2 GI:13757702
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 594)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
BDGP/HMI Drosophila EST Project
Unpublished (2001)
JOURNAL On Aug 26, 1998 this sequence version replaced gi:3478104.
COMMENT Other_ESTs: GH09125.3prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003684: arm:3R [5513047,5732625]
estimated cyto:85E6-85E2: 04/10/2001
Plate: GH.91 row: C column: 1
High quality sequence stop: 503
POLYA-No.
FEATURES
Source
1..594
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH09125"
/clone_lib="GH Drosophila melanogaster head pOT2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT 181 a 129 c 145 g 139 t
ORIGIN
Query Match 80.0%; Score 16.8; DB 10; Length 594;
Best Local Similarity 90.0%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 2 taacatagatgacaccgcgc 21
Db 273 TCACATAGATGACACCGCGC 292

```

```

RESULT 11
AZ029768      662 bp  DNA      GSS      25-FEB-2000
LOCUS      RPCI-23-349G17.TV RPCI-23 Mus musculus genomic clone RPCI-23-349G17
DEFINITION      DNA sequence.
ACCESSION      AZ029768
VERSION      AZ029768.1  GI:7105152
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 662)
AUTHORS      Zhao,S., Nierman,W., Feidblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL      Unpublished (1999)
COMMENT      Other_GSSs: RPCI-23-349G17.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 349 row: G column: 17
Seq primer: T7
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..662
     /organism="Mus musculus"
     /strain="C57BL/6J"
     /db_xref="taxon:10090"
     /clone="RPCI-23-349G17"
     /clone_lib="RPCI-23"
     /sex="Female"
     /lab_host="DH10B"
     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
     brain genomic DNA was isolated and partially digested
     with a combination of EcoRI and EcoRI Methylase. Size
     selected DNA was cloned into the pBACe3.6 vector at the
     EcoRI sites. The ligation products were transformed into
     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      180 a 136 c 175 g 170 t
ORIGIN
1 gtaacatagatgacaccgcg 21
|||||
Db 143 TAAGATAGATGACACAGCGC 162

Query Match      80.0%; Score 16.8; DB 13; Length 662;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 12
BE038757      710 bp  mRNA      EST      07-JUN-2000
LOCUS      BE038757
DEFINITION      AB060007 AB Arabidopsis thaliana cDNA 5' similar to dna-binding
protein, mRNA sequence.
ACCESSION      BE038757
VERSION      BE038757.1  GI:8333878
KEYWORDS      EST.
FEATURES             Location/Qualifiers
     source           1..826
     /organism="Arabidopsis thaliana"
     /cultivar="Columbia"
     /db_xref="taxon:3702"
     /clone_lib="AB"
     /tissue_type="leaves, seedlings"
     /dev_stage="2-3 weeks"
     /note="200mM NaCl"
BASE COUNT      255 a 135 c 177 g 143 t
ORIGIN
1 gtaacatagatgacaccgcg 20
|||||
Db 160 GTAAACAAGATAACACCGCG 179

Query Match      80.0%; Score 16.8; DB 10; Length 710;
Best Local Similarity 90.0%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 13
AW940028      826 bp  mRNA      EST      19-APR-2001
LOCUS      GH01164.3prime GH Drosophila melanogaster head p072 Drosophila
DEFINITION      melanogaster cDNA clone GH01164 3, mRNA sequence.
ACCESSION      AW940028
VERSION      AW940028.1  GI:8115482
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 826)
AUTHORS      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Other_ESTs: GH01164.5prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or
more T residues at the beginning of the sequence, this clone was
polyadenylated. The resulting poly-T sequence has been removed.
Plate: GH.11 row: F column: 4
High quality sequence stop: 795.
FEATURES             Location/Qualifiers
     source           1..826

```

```

SOURCE      thale cress.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE      1 (bases 1 to 710)
AUTHORS      Bohnert,H.J., Borchert,C., Brazillie,S., Brooks,J., Eaton,M., Ferrea
,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
Scara,G., Wheeler,M. and Zepeda,G.R.
TITLE      Functional Genomics of Plant Stress Tolerance
JOURNAL      Unpublished (2000)
COMMENT      Contact: Michalowski,C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu
An open reading frame exists.
Insert Length: 1 Std Error: 0.00.
FEATURES             Location/Qualifiers
     source           1..710
     /organism="Arabidopsis thaliana"
     /cultivar="Columbia"
     /db_xref="taxon:3702"
     /clone_lib="AB"
     /tissue_type="leaves, seedlings"
     /dev_stage="2-3 weeks"
     /note="200mM NaCl"
BASE COUNT      255 a 135 c 177 g 143 t
ORIGIN
1 gtaacatagatgacaccgcg 20
|||||
Db 160 GTAAACAAGATAACACCGCG 179

Query Match      80.0%; Score 16.8; DB 10; Length 710;
Best Local Similarity 90.0%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 13
AW940028      826 bp  mRNA      EST      19-APR-2001
LOCUS      GH01164.3prime GH Drosophila melanogaster head p072 Drosophila
DEFINITION      melanogaster cDNA clone GH01164 3, mRNA sequence.
ACCESSION      AW940028
VERSION      AW940028.1  GI:8115482
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 826)
AUTHORS      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Other_ESTs: GH01164.5prime
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Based upon the presence of a XhoI site followed by a run of 14 or
more T residues at the beginning of the sequence, this clone was
polyadenylated. The resulting poly-T sequence has been removed.
Plate: GH.11 row: F column: 4
High quality sequence stop: 795.
FEATURES             Location/Qualifiers
     source           1..826

```



Search completed: February 25, 2002, 17:20:51  
Job time: 16144 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: February 25, 2002, 17:59:54 ; Search time 2331.3 Seconds  
(without alignments)  
148.604 Million cell updates/sec

Title: US-09-698-903B-3  
Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues  
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_on.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	100.0	21	6	AX127750	AX127750 Sequence
2	21	100.0	3266	6	A08065	A08065 Synthetic n
3	21	100.0	3266	6	A10061	A10061 N.tabacum (
c 4	21	100.0	4832	6	AX172441	AX172441 Sequence
c 5	21	100.0	4946	6	A60108	A60108 Sequence 1
c 6	21	100.0	4946	6	A76915	A76915 Sequence 1
c 7	21	100.0	4946	6	AR098307	AR098307 Sequence
c 8	21	100.0	4946	6	AX172440	AX172440 Sequence
c 9	21	100.0	5865	6	AX127748	AX127748 Sequence
c 10	21	100.0	6254	8	NTA29	X52283 Tobacco ant
c 11	19.4	92.4	3243	5	XL070985	U70985 Xenopus lae
c 12	19.4	92.4	149710	2	AC022563	AC022563 Homo sapi
c 13	19.4	92.4	179183	2	AC019254	AC019254 Homo sapi
c 14	17.8	84.8	167409	2	AC073374	AC073374 Homo sapi
c 15	17.8	84.8	179510	2	AC013820	AC013820 Homo sapi
c 16	17.4	82.9	39443	9	AC010505	AC010505 Homo sapi
c 17	17.4	82.9	172246	2	AC024941	AC024941 Homo sapi
c 18	17.4	82.9	195516	2	AC053540	AC053540 Homo sapi
c 19	17.4	82.9	207856	2	AC078778	AC078778 Homo sapi
c 20	17	81.0	4900	5	XLAPICAL	Z14997 X.laavis mr
c 21	17	81.0	123016	2	AF189001	AF189001 Homo sapi
c 22	17	81.0	159934	9	AL160273	AL160273 Human DNA
c 23	17	81.0	182942	2	AC087793	AC087793 Homo sapi
c 24	17	81.0	186413	2	AL161913	AL161913 Homo sapi
c 25	16.8	80.0	1485	14	AF195032	AF195032 Dengue vi
c 26	16.8	80.0	1485	14	AF195033	AF195033 Dengue vi
c 27	16.8	80.0	1485	14	AF195034	AF195034 Dengue vi
c 28	16.8	80.0	1485	14	AF195035	AF195035 Dengue vi
c 29	16.8	80.0	1485	14	AF195036	AF195036 Dengue vi
c 30	16.8	80.0	1485	14	AF195037	AF195037 Dengue vi
c 31	16.8	80.0	1485	14	AF195038	AF195038 Dengue vi
c 32	16.8	80.0	1485	14	AF195039	AF195039 Dengue vi
c 33	16.8	80.0	1485	14	AF264053	AF264053 Dengue vi
c 34	16.8	80.0	1485	14	DENENVB	M24445 Dengue viru
c 35	16.8	80.0	1485	14	DENENVC	M24447 Dengue viru
c 36	16.8	80.0	1485	14	DENENVD	M24448 Dengue viru
c 37	16.8	80.0	1485	14	DVU34933	U34933 Dengue viru
c 38	16.8	80.0	1485	14	DVU34934	U34934 Dengue viru
c 39	16.8	80.0	1485	14	DVU34935	U34935 Dengue viru
c 40	16.8	80.0	1485	14	DVU34936	U34936 Dengue viru
c 41	16.8	80.0	1485	14	DVU34937	U34937 Dengue viru
c 42	16.8	80.0	1485	14	DVU34938	U34938 Dengue viru
c 43	16.8	80.0	1485	14	DVU34939	U34939 Dengue viru
c 44	16.8	80.0	1485	14	DVU34940	U34940 Dengue viru
c 45	16.8	80.0	1485	14	TOGDENW2	X15433 Dengue viru

ALIGNMENTS

RESULT 1	AX127750	AX127750	21 bp	DNA	PAT	15-MAY-2001
LOCUS	Sequence 3	from Patent WO0131042.				
DEFINITION	AX127750					
ACCESSION	AX127750.1	GI:14134397				
VERSION						
KEYWORDS	synthetic construct.					
SOURCE	synthetic construct					
ORGANISM	artificial sequence.					
REFERENCE	1 (bases 1 to 21)					
AUTHORS	Weston,B. and de Beuckeleer,M.					
TITLE	Male-sterile brassica plants and methods for producing same					
JOURNAL	Patent: WO 0131042-A 3 03-MAY-2001;					
FEATURES	Aventis CropScience N.V. (BE)					
source	Location/Qualifiers					
	1..21					
	/organism="synthetic construct"					
	/db_xref="taxon:32630"					
	/note="primer MLD008"					

```
BASE COUNT      4 a      1 c      10 g      6 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
|||||
Db 1 ATAGGTGGGAGGCTATTGG 21

RESULT 2
A08065      A08065      3266 bp      DNA      PAT      12-AUG-1993
LOCUS      Synthetic nucleotide sequence of the TA29 gene.
DEFINITION
ACCESSION  A08065
VERSION     A08065.1 GI:413313
KEYWORDS   .
SOURCE     synthetic construct.
ORGANISM   synthetic construct.
REFERENCE  1 (bases 1 to 3266)
AUTHORS    .
JOURNAL    Patent: WO 8910396-A 7 02-NOV-1989;
FEATURES   Location/Qualifiers
            source          1..3266
                        /organism="synthetic construct"
                        /db_xref="taxon:32630"
                        1527..2492
                        /gene="TA29"
                        1527..2492
                        /gene="TA29"
                        /codon_start=1
                        /transl_table=1
                        /protein_id="CAA00738.1"
                        /db_xref="GI:413314"
                        /translation="MVAPKWFISFMILLSLAICSGQPVTSDAIKAKADHNLKAHT
                        LSNIDAKFGGGGGGIGGAGGAGGGGGSDTPNNGYNPGCSIHGCTVPGGFL
                        PKPVFGVPVSPGCGYCPADIPGTGMPTESKITGISQARLYRCAPGNMCDKDCNE
                        LLLHFVPMQDKHDKQHLRYGGRGIGLTVGGVGGFIGFGAMGGGGGGGGSDA
                        PGCSNDGDPGCGPGCGYACPANNPGSGGITFEFHSLSRFDGYPYRCRPMCESEDC
                        NELLHFVSPMQHKHNRHDIHVERSDDEEAHHQSKQKHDEDIIN"
BASE COUNT      1016 a      581 c      623 g      1046 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 3266;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
|||||
Db 178 ATAGGTGGGAGGCTATTGG 198

RESULT 4
AX172441/c      AX172441      4832 bp      DNA      PAT      03-JUL-2001
LOCUS      Sequence 2 from Patent WO0141558.
DEFINITION  AX172441
ACCESSION   AX172441
VERSION     AX172441.1 GI:14597553
KEYWORDS   .
SOURCE     synthetic construct.
ORGANISM   synthetic construct.
REFERENCE  1 (bases 1 to 4832)
AUTHORS    de Both, G. and de Beuckeleer, M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL    Patent: WO 0141558-A 2 14-JUN-2001;
           Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source          1..4832
                        /organism="synthetic construct"
                        /db_xref="taxon:32630"
                        /note="n-DNA of plasmid pTHW118"
                        1883..4065
                        /note="HpaI restriction fragment"
BASE COUNT      1528 a      883 c      932 g      1488 t
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
|||||
Db 4602 ATAGGTGGGAGGCTATTGG 4582

RESULT 5
A60108/c      A60108      4946 bp      DNA      PAT      06-MAR-1998
LOCUS      Sequence 1 from Patent WO9706267.
DEFINITION  A60108
ACCESSION   A60108
VERSION     A60108.1 GI:3715124
KEYWORDS   .
SOURCE     common tobacco.
ORGANISM   Nicotiana tabacum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
           1 (bases 1 to 3266)
           Mariani, C., Leemans, J., De Greef, W. and De Beuckeleer, M.
           Plants with modified stamen cells
           Patent: EP 0344029-A 9 29-NOV-1989;
           100.0%; Score 21; DB 6; Length 3266;
           Best Local Similarity 100.0%; Pred. No. 0.44;
           Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
|||||
Db 178 ATAGGTGGGAGGCTATTGG 198

RESULT 3
A10061      A10061      3266 bp      DNA      PAT      07-FEB-1994
LOCUS      N.tabacum (pTA2953) TA29 gene.
DEFINITION  A10061
ACCESSION   A10061
VERSION     A10061.1 GI:490021
KEYWORDS   .
SOURCE     common tobacco.
ORGANISM   Nicotiana tabacum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
           1 (bases 1 to 3266)
           Mariani, C., Leemans, J., De Greef, W. and De Beuckeleer, M.
           Plants with modified stamen cells
           Patent: EP 0344029-A 9 29-NOV-1989;
           100.0%; Score 21; DB 6; Length 3266;
           Best Local Similarity 100.0%; Pred. No. 0.44;
           Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
|||||
Db 178 ATAGGTGGGAGGCTATTGG 198
```

## KEYWORDS

SOURCE Transformation vector pTHW107.  
ORGANISM Transformation vector pTHW107  
artificial sequence; vectors.

REFERENCE 1 (bases 1 to 4946)

AUTHORS De,B.M.

TITLE Genetic transformation using a PARP inhibitor

JOURNAL Patent: WO 9706267-A 1 20-FEB-1997;

PLANT GENETIC SYSTEMS NV (BE)

Location/Qualifiers

FEATURES

source 1..4946

/organism="transformation vector pTHW107"

/db\_xref="taxon:126810"

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 6

A76915/c

LOCUS A76915 4946 bp DNA

DEFINITION Sequence 1 from Patent EP0757102.

ACCESSION A76915

VERSION A76915.1 GI:6088712

KEYWORDS Transformation vector pTHW107.

SOURCE Transformation vector pTHW107

ORGANISM artificial sequence; vectors.

REFERENCE 1 (bases 1 to 4946)

AUTHORS De,B.M.

TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR

JOURNAL Patent: EP 0757102-A 1 05-FEB-1997;

PLANT GENETIC SYSTEMS NV (BE)

Location/Qualifiers

FEATURES

source 1..4946

/organism="Transformation vector pTHW107"

/db\_xref="taxon:126810"

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 7

AR098307/c

LOCUS AR098307 4946 bp DNA

DEFINITION Sequence 1 from patent US 6074876.

ACCESSION AR098307

VERSION AR098307.1 GI:12807564

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 4946)

AUTHORS De Block,M.

TITLE Genetic transformation using a PARP inhibitor

JOURNAL Patent: US 6074876-A 1 13-JUN-2000;

Location/Qualifiers

FEATURES

source

1..4946

/organism="unknown"

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 8

AX172440/c

LOCUS AX172440 4946 bp DNA

DEFINITION Sequence 1 from Patent WO0141558.

ACCESSION AX172440

VERSION AX172440.1 GI:14597552

KEYWORDS synthetic construct.

SOURCE synthetic construct

ORGANISM artificial sequence.

REFERENCE 1 (bases 1 to 4946)

AUTHORS de Both,G. and de Beuckeleer,M.

TITLE Hybrid winter oilseed rape and methods for producing same

JOURNAL Patent: WO 0141558-A 1 14-JUN-2001;

Aventis CropScience N.V. (BE)

Location/Qualifiers

FEATURES

source 1..4946

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="T-DNA of plasmid pTHW107"

misc\_feature 964..4906

/note="Hind III fragment"

BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 4946;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 9

AX127748/c

LOCUS AX127748 5865 bp DNA

DEFINITION Sequence 1 from Patent WO0131042.

ACCESSION AX127748

VERSION AX127748.1 GI:14134395

KEYWORDS synthetic construct.

SOURCE synthetic construct

ORGANISM artificial sequence.

REFERENCE 1 (bases 1 to 5865)

AUTHORS Weston,B. and de Beuckeleer,M.

TITLE Male-sterile brassica plants and methods for producing same

JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;

Aventis CropScience N.V. (BE)

Location/Qualifiers

FEATURES

source 1..5865

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="T-DNA of plasmid pCOL13"

BASE COUNT 1849 a 1095 c 1149 g 1772 t

ORIGIN

Query Match	100.0%;	Score 21;	DB 6;	Length 5865;
Best Local Similarity	100.0%;	Pred. No. 0.46;		
Matches	21;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	atagggtggaggctattgg	21	
Db	4717	ATAGGGTGGAGGCTATTGG	4697	
RESULT	10			
LOCUS	NPTA29	6254 bp	DNA	PLN
DEFINITION	Tobacco anther-specific gene TA-29 and stem-specific gene TSJt1.			
ACCESSION	X52283			
VERSION	X52283.1 GI:20035			
KEYWORDS	tissue specific gene.			
SOURCE	common tobacco.			
ORGANISM	Nicotiana tabacum			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.			
AUTHORS	Mariani, T.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-MAR-1990) Mariani T., Plant Genetic Systems, J Plateaustraat 22, B 9000 Gent, Belgium			
REFERENCE	2 (bases 1 to 6254)			
AUTHORS	Seurinck, J., Truettner, J. and Goldberg, R.B.			
TITLE	The nucleotide sequence of an anther-specific gene			
JOURNAL	Nucleic Acids Res. 18 (11), 3403 (1990)			
MEDLINE	9028727			
COMMENT	Data kindly reviewed (13-DEC-1990) by Mariani C.			
FEATURES	Location/Qualifiers			
source	1..6254			
	/organism="Nicotiana tabacum"			
	/cultivar="Samsun"			
	/db_xref="taxon:4097"			
	/clone_lib="lambda Charon32"			
TATA_signal	1446..1452			
	/note="TSJt1"			
gene	1477..2714			
	/gene="TA29"			
prim_transcript	1477..(2617,2714)			
	/gene="TA29"			
	/note="anther-specific"			
CDS	1527..2492			
	/gene="TA29"			
	/note="anther-specific"			
	/codon_start=1			
	/protein_id="CAA36524.1"			
	/db_xref="GI:20036"			
	/db_xref="SWISS-PROT:P24804"			
	/translation="MVAPKWFISFMILLISLAICSGQPVTSDAIRAKEADHNLKAHT LSNDAFGGFGGGGGGGSDPTNYGPNPSCIHGCTVPGGFL PKNDFGVPSGCGVYCPADPTGWTESKITIGISQARLYRCKPGPMCDKDCNE LLHFVPMQDKNDKQHEHLRYGRRIGLTVGGVFGIFGAMGGGGGGGSDA PGCSNDCDFGCGPCGCVACPNPNPSSGIGIEFISGLSRFDGYPYRCPDCESEDN NELLHVPSPMOKHENRHDHIVERDEAHQSKQKDEIIN"			
mat_peptide	1527..2489			
	/gene="TA29"			
	/product="anther-specific"			
prim_transcript	complement(<3022..>4678)			
	/gene="TSJt1"			
exon	/note="stem specific, weakly expressed in other organs"			
	complement(<3022..3075)			
	/gene="TSJt1"			
	/number=3			
mRNA	complement(join(<3022..3075,4047..4287,4473..>4678))			
	/gene="TSJt1"			
gene	/note="stem specific, weakly expressed in other organs"			
	complement(3022..4678)			
	/gene="TSJt1"			



```

exon      <1..3224
          /number=1
CDS       2891..>3224
          /codon_start=1
          /product="transforming growth factor-beta 5 (TGF-beta 5)"
          /protein_id="AAC60371.1"
          /db_xref="GI:2281072"
          /translation="MEVLNMLLVLLVLLHSSLSMSLSTCKAVDMEEVRRKRRIEATRGQ
ILSKLKDTPDVSSEKMPVPEAFLYNSTLEIVIREKATREEEHVGHQDNIQDYAK
QYRFESIT"
intron    3225..>3243
          /number=1
BASE COUNT      970 a   675 c   676 g   922 t
ORIGIN
Query Match      92.4%;   Score 19.4;   DB 5;   Length 3243;
Best Local Similarity 95.2%;   Pred.No. 3.6;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  1 atagggtggaggcattgtg 21
      |||||
Db  1731 ATAGGTGGGAGGCTATGG 1711

RESULT 12
AC022563/c
LOCUS      AC022563      149710 bp      DNA      HTG      13-JUL-2000
DEFINITION Homo sapiens clone Rp11-3L23, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC022563
VERSION   AC022563.2  GI:9121084
KEYWORDS  HTG; HTGS_PHASE0.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149710)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens chromosome, clone Rp11-3L23
JOURNAL   unpublished
REFERENCE 2 (bases 1 to 149710)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
           Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
           Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
           DeAtrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
           Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galegan,J.,
           Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
           Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
           Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
           McPheeters,R., Meldrin,J., Meneus,L., Morrow,J., Naylor,J.,
           Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
           Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
           Roy,A., Santos,R., Severi,P., Spencer,B., Stange-Thomann,N.,
           Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
           Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6910806.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
           Center project name: L2771
           Center clone name: 3.L.23

```





\* 146081 163677: contig of 17597 bp in length  
 \* 163678 163777: gap of 100 bp  
 \* 163778 167409: contig of 3632 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1. 167409  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-44108"  
 /clone\_lib="RPC1-11 Human Male BAC"

misc\_feature  
 1. 11391  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"

misc\_feature  
 11492. 113778  
 /note="assembly\_fragment"

misc\_feature  
 13879. 15678  
 /note="assembly\_fragment"

misc\_feature  
 15779. 17278  
 /note="assembly\_fragment"

misc\_feature  
 17379. 19365  
 /note="assembly\_fragment"

misc\_feature  
 19466. 22846  
 /note="assembly\_fragment"

misc\_feature  
 22947. 26400  
 /note="assembly\_fragment"

misc\_feature  
 26501. 28910  
 /note="assembly\_fragment"

misc\_feature  
 29011. 51360  
 /note="assembly\_fragment"

misc\_feature  
 51461. 55176  
 /note="assembly\_fragment"

misc\_feature  
 55277. 58304  
 /note="assembly\_fragment"

misc\_feature  
 58405. 62806  
 /note="assembly\_fragment"

misc\_feature  
 62907. 65285  
 /note="assembly\_fragment"

misc\_feature  
 69386. 79212  
 /note="assembly\_fragment"

misc\_feature  
 79313. 87095  
 /note="assembly\_fragment"

misc\_feature  
 87196. 95331  
 /note="assembly\_fragment"

misc\_feature  
 93432. 102696  
 /note="assembly\_fragment"

misc\_feature  
 102797. 111775  
 /note="assembly\_fragment"

misc\_feature  
 111876. 122023  
 /note="assembly\_fragment"

misc\_feature  
 122124. 132655  
 /note="assembly\_fragment"

misc\_feature  
 132756. 145980  
 /note="assembly\_fragment"

misc\_feature  
 146081. 163677  
 /note="assembly\_fragment"

misc\_feature  
 163778. 167409  
 /note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:right"

BASE COUNT 51934 a 30270 c 30508 g 52497 t 2200 others  
 ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 167409;  
 Best Local Similarity 90.5%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21  
 ||||| ||||| ||||| |||||

Db 126062 ATAGAGTTGGAGGCTATTTCG 126042

## RESULT 15

AC013820  
 LOCUS  
 DEFINITION  
 AC013820 179510 bp DNA HTG 01-APR-2000  
 Homo sapiens clone RP11-21P24, WORKING DRAFT SEQUENCE, 8 unordered  
 pieces.  
 AC013820  
 VERSION  
 AC013820.3 GI:7382157  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 179510)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens, clone RP11-21P24  
 Unpublished  
 2 (bases 1 to 179510)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
 Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,  
 Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,  
 Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,  
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,  
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,  
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 1, 2000 this sequence version replaced gi:6573887.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L4081  
 Center clone name: 21\_P24  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 175240 bases at least Q40  
 Consensus quality: 176804 bases at least Q30  
 Consensus quality: 177720 bases at least Q20  
 Insert size: 183000; agarose-fp  
 Insert size: 178810; sum-of-contigs  
 Quality coverage: 6.8 in Q20 bases; agarose-fp  
 Quality coverage: 7.0 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 2601: contig of 2601 bp in length  
 \* 2602 2701: gap of 100 bp  
 \* 2702 8297: contig of 5596 bp in length  
 \* 8298 8397: gap of 100 bp  
 \* 8398 27253: contig of 18856 bp in length  
 \* 27254 27353: gap of 100 bp

TITLE  
JOURNAL

## COMMENT

\* 27354 41791: contig of 14438 bp in length  
\* 41792 41891: gap of 100 bp  
\* 41892 61037: contig of 19146 bp in length  
\* 61038 61137: gap of 100 bp  
\* 61138 88113: contig of 28976 bp in length  
\* 88114 88213: gap of 100 bp  
\* 88214 131619: contig of 43406 bp in length  
\* 131620 131719: gap of 100 bp  
\* 131720 179510: contig of 47791 bp in length.

FEATURES

source

1..179510  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-21P24"  
/clone\_lib="RPC1-11 Human Male BAC"

misc\_feature

1..2601  
/note="assembly\_fragment"

misc\_feature

2702..8297  
/note="assembly\_fragment"

misc\_feature

8398..27253  
/note="assembly\_fragment"

misc\_feature

27354..41791  
/note="assembly\_fragment"

clone\_end:SP6

vector\_side:right

misc\_feature

41892..61037  
/note="assembly\_fragment"

misc\_feature

61138..88113  
/note="assembly\_fragment"

misc\_feature

88214..131619  
/note="assembly\_fragment"

clone\_end:T7

vector\_side:right

misc\_feature

131720..179510  
/note="assembly\_fragment"

BASE COUNT 59127 a 32374 c 31805 g 55501 t 703 others

ORIGIN

Query Match 84.8%; Score 17.8; DB 2; Length 179510;  
Best Local Similarity 90.5%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 atagggtggaggctatttgg 21  
|||||  
Db 107231 ATAGAGTTGGAGGCTATTGG 107251

Search completed: February 25, 2002, 18:00:06  
Job time: 18414 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:19 ; Search time 716.55 Seconds  
(without alignments)  
25.126 Million cell updates/sec

Title: US-09-698-903B-3  
Perfect score: 21  
Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	21	100.0	21	22	PCR primer MLD008
2	21	100.0	3265	10	TA29 gene insert 1
3	21	100.0	4832	22	Nucleotide sequenc
4	21	100.0	4946	18	T-DNA of plasmid p
5	21	100.0	4946	22	Nucleotide sequenc
6	21	100.0	5864	17	Plasmid pTCO113 T-
7	21	100.0	5865	22	Chimeric T-DNA of
8	16.8	80.0	1983	22	Dengue-2 virus der
9	16.8	80.0	10723	18	cDNA encoding poly
10	16.8	80.0	10723	18	cDNA sequence enco
11	16.4	78.1	694	22	Human lung tumour

c 12	16.4	78.1	857	14	AAQ46850	Recombinant human
c 13	16.4	78.1	857	17	AAQ46850	Human milk kappa-c
c 14	16.4	78.1	858	16	AAQ89598	Human kappa-casein
c 15	16.4	78.1	2185	20	AAZ21101	Human secreted pro
c 16	16.4	78.1	13104	14	AAQ46852	Clone of recombina
c 17	16.2	77.1	1418	20	AAQ4867	rodent DCMPI C-lec
c 18	16.2	77.1	24978	20	AAQ60209	Genomic DNA sequen
c 19	16.2	77.1	24979	21	AAQ60209	SEQ ID 3 of US5914
c 20	15.8	75.2	394	21	AAQ3474	Human secreted pro
c 21	15.8	75.2	660	21	AAQ12948	Aspergillus oryzae
c 22	15.8	75.2	777	15	AAQ70891	Tomato spotted wil
c 23	15.8	75.2	1026	16	AAQ94047	Human ALD gene exo
c 24	15.8	75.2	1549	22	AAH14886	Human cDNA sequenc
c 25	15.8	75.2	1556	22	AAH74185	s1 RNA binding reg
c 26	15.8	75.2	1577	22	AAI59214	Human polynucleoti
c 27	15.8	75.2	1591	22	AAH16816	Human cDNA sequenc
c 28	15.8	75.2	1602	22	AAH15727	Human cDNA sequenc
c 29	15.8	75.2	1643	19	AAV43711	Human sodium-depen
c 30	15.8	75.2	1644	21	AAQ76111	Human ORFX ORF1666
c 31	15.8	75.2	1795	19	AAV57910	Human haemochromat
c 32	15.8	75.2	3049	15	AAQ70890	Tomato spotted wil
c 33	15.8	75.2	9139	21	AAZ35274	Soybean retroelime
c 34	15.4	73.3	2338	22	AAH16691	Human cDNA sequenc
c 35	15.4	73.3	3024	20	AAZ42247	Human normal bladd
c 36	15.4	73.3	4636	19	AAV23920	Human alpha3 integ
c 37	15.4	73.3	5769	19	AAV15004	Receptor protein t
c 38	15.4	73.3	44453	20	AAZ3519	Human kidney amino
c 39	15.2	72.4	139	12	AAQ13998	Template DNA. Syn
c 40	15.2	72.4	140	14	AAQ35327	DNA amplification
c 41	15.2	72.4	455	22	AAI12272	Probe #2205 for ge
c 42	15.2	72.4	455	22	AAI33627	Probe #2313 used t
c 43	15.2	72.4	455	22	AAI02189	Probe #2180 used t
c 44	15.2	72.4	649	21	AAI14117	Aspergillus oryzae
c 45	15.2	72.4	1302	21	AAQ50359	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1  
AAD06992  
ID AAD06992 standard; DNA; 21 BP.  
XX  
AC AAD06992;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE PCR primer MLD008 to amplify a fragment of pTCO113.  
XX  
KW plasmid pTCO113; transgenic Brassica plant; transformation event;  
KW male-sterility gene; tobacco; PCR primer; probe; ss.  
XX  
OS Chimeric - Bacillus amyloliquefaciens.  
OS Chimeric - Nicotiana tabacum.  
XX  
PN WO200131042-A2.  
XX  
PD 03-MAY-2001.  
XX  
PF 26-OCT-2000; 2000WO-EP10680.  
XX  
PR 29-OCT-1999; 99US-0430497.  
XX  
PA (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
PI Weston B, De Beuckeleer M;  
XX  
DR WPI; 2001-300517/31.  
XX  
PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX

```

PS Claim 1; Page 26; 53pp; English.
XX
CC The present invention relates to a transgenic Brassica plant or its
CC seed, cells or tissues, characterised by harbouring a specific
CC transformation event, particularly by the presence of a male-sterility
CC gene, at a specific location in the Brassica genome. Transgenic
CC Brassica plant is useful for producing a hybrid seed by crossing the
CC transgenic plant with a male-fertile Brassica plant and harvesting the
CC hybrid seed from the transgenic Brassica plant.
CC The present sequence is PCR primer MLD008 used to amplify a fragment of
CC plasmid pTCO113. The primer is also used as a probe. The amplified
CC fragment comprises coding region of barnase from
CC Bacillus amyloliquefaciens and promoter region of the anther-specific
CC gene TA29 from Nicotiana tabacum. This primer corresponds to position
CC 4697-4717 of plasmid pTCO113.
XX
SQ Sequence 21 BP; 4 A; 1 C; 10 G; 6 T; 0 other;

Query Match      100.0%; Score 21; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggcatttgg 21
Db 1 atagggtggaggcatttgg 21

RESULT 2
AAN92188
ID AAN92188 standard; DNA; 3265 BP.
XX
AC AAN92188;
XX
DT 11-APR-1990 (first entry)
XX
DE TA29 gene insert in pTA29S3.
XX
KW TA29 promoter; pTA29S3;
XX
OS Nicotiana tabacum "Samsun".
XX
FH Key Location/Qualifiers
FT TATA-signal 1445..1451
FT /tag= a
FT CDS 1526..4991
FT /tag= b
FT CDS 1940..2296
FT /tag= c
XX
PN WO8910396-A.
XX
XX 02-NOV-1989.
XX
XX 27-APR-1989; 89WO-EP00495.
XX
XX 28-APR-1988; 88GB-0010120.
XX
XX (PLAN-) PLANT GENETIC SYSTEMS NV.
XX
XX Mariani C, Leemans J, De Greef W, De Beuckeleer M;
XX WPI: 1989-339961/46.
XX P-PSDB; AAP93313.
XX
XX Transformed plant cell with disrupted metabolism - from a product encoded
XX by foreign male-sterility DNA inserted in the nuclear genome.
XX
XX Example 2; fig. 3A; 89pp; English.
XX
XX Plasmid pTA29S3 was constructed subcloning a ClaI fragment contg. the
XX TA29 gene from lambda TA29 into pLX31. The TA29 promoter is used in a
XX vector for transforming plant cells. Tag b is the coding sequence of the

```

```

CC TA29 gene; tag b is the cDNA insert from pBR329 (AAN92187).
XX
SQ Sequence 3265 BP; 1016 A; 581 C; 623 G; 1045 T; 0 other;

Query Match      100.0%; Score 21; DB 10; Length 3265;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggcatttgg 21
Db 177 atagggtggaggcatttgg 197

RESULT 3
AAH25423/C
ID AAH25423 standard; DNA; 4832 BP.
XX
AC AAH25423;
XX
DT 22-AUG-2001 (first entry)
XX
DE Nucleotide sequence of plasmid pTHW118.
XX
KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barstar gene; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
OS Arabidopsis thaliana.
OS Bacillus amyloliquefaciens.
OS Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "right border repeat from TL-DNA from pTiB6S3"
FT misc_feature 26..53
FT /tag= b
FT /note= "synthetic polylinker derived sequences"
FT misc_feature 54..90
FT /tag= c
FT /note= "residual sequence from TL-DNA at right border repeat"
FT misc_feature 91..97
FT /tag= d
FT /note= "synthetic polylinker derived sequences"
FT 3'UTR complement (98..309)
FT /tag= e
FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"
FT misc_feature 310..330
FT /tag= f
FT /note= "synthetic polylinker derived sequences"
FT CDS complement (331..882)
FT /tag= g
FT /note= "Streptomyces hygroscopicus bialaphos resistance (bar) gene"
FT promoter complement (883..2608)
FT /tag= h
FT /note= "atsIA ribulose-1,5-bisphosphate carboxylase small subunit gene from Arabidopsis thaliana"
FT misc_feature 2609..2658
FT /tag= i
FT /note= "synthetic polylinker derived sequences"
FT 3'UTR complement (2659..2919)
FT /tag= j
FT /note= "taqI fragment from 3' UTR of nopaline synthase gene from T-DNA of pTi37 and containing plant polyadenylation signals"
FT misc_feature 2920..2940
FT /tag= k
FT /note= "synthetic polylinker derived sequences"
FT 3'UTR 2941..2980

```





CC pTHW107 and helper T1 plasmid pMP60. In some treatments, the  
 CC hypocotyls were treated with the poly-(ADP-ribose) polymerase  
 CC inhibitor niacinamide (250 mg/l) 4 days prior to infection.  
 CC Plants regenerated from niacinamide-treated transformed calli  
 CC had a low copy number and displayed less variation in the  
 CC expression profile of the transgenes.  
 XX  
 SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 100.0%; Score 21; DB 18; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.42; Mismatches 0; Indels 0; Gaps 0;

OY 1 atagggtggaggctatttgg 21  
 |||||  
 Db 4716 ATAGGTTGGAGGCTATTGG 4696

RESULT 5  
 AAH25422/c  
 ID AAH25422 standard; DNA; 4946 BP.

XX AC AAH25422;

XX DT 22-AUG-2001 (first entry)

XX Nucleotide sequence of plasmid pTHW107.

DE Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barnase gene; ss.

XX Synthetic.

OS Streptomyces hygroscopicus.

OS Arabidopsis thaliana.

OS Bacillus amyloliquefaciens.

XX Nicotiana tabacum.

FH Key Location/Qualifiers

FT misc\_feature 1..25

FT /tag= a

FT /note= "right border repeat from TL-DNA from pTiB6S3"

FT misc\_feature 26..97

FT /tag= b

FT /note= "synthetic polylinker derived sequences"

FT complement (98..309)

FT /tag= c

FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"

FT misc\_feature 310..330

FT /tag= d

FT /note= "synthetic polylinker derived sequences"

FT CDS 331..882

FT /tag= e

FT /note= "Streptomyces hygroscopicus bar gene"

FT complement (883..2608)

FT /tag= f

FT /note= "AtSIA ribulose-1,5-biphosphate carboxylase

FT misc\_feature 2609..2658

FT /tag= g

FT /note= "synthetic polylinker derived sequences"

FT complement (2659..2919)

FT /tag= h

FT /note= "Taql fragment from 3' UTR of nopaline

FT synthase gene from T-DNA of pTi37 and

FT containing plant polyadenylation signals"

FT 3'UTR 2920..3031

FT /tag= i

FT /note= "3'UTR downstream of Bacillus amyloliquefaciens

FT barnase coding region"

FT CDS complement (3032..3367)

FT /tag= j

FT /note= "Barnase coding region from Bacillus

FT promoter complement amyloliquefaciens"  
 FT /tag= k  
 FT /note= "anther-specific gene TA29 promoter from  
 FT Nicotiana tabacum"  
 FT misc\_feature 4878..4921  
 FT /tag= l  
 FT /note= "synthetic polylinker derived sequences"  
 FT misc\_feature 4922..4946  
 FT /tag= m  
 FT /note= "left border repeat from TL-DNA from pTiB6S3"

WO200141558-A1.

14-JUN-2001.

06-DEC-2000; 2000WO-EP12872.

08-DEC-1999; 99US-0457037.

(AVET ) AVENTIS CROPS SCIENCE NV.

De Both G, De Beuckeleer M;

WPI; 2001-381419/40.

Transgenic winter oilseed rape plants suited for producing hybrid seed  
 with improved qualities, comprises a male-sterility gene and fertility  
 restorer gene, integrated into the genome -

Example 1; Page 78-80; 98pp; English.

The specification describes a pair of transgenic winter oilseed rape  
 plants suited for producing hybrid seed. One of the plants has an  
 expression cassette comprising a male-sterility gene, and the other  
 plant has an expression cassette comprising a fertility restorer gene,  
 integrated into the genome. The fertility restorer gene is capable of  
 preventing the activity of the male-sterility gene. The plant pair is  
 useful for producing hybrid seed. Plants developed from the hybrid  
 seed have agronomic performance, genetic stability and adaptability to  
 different genetic backgrounds. The present sequence represents  
 plasmid pTHW107. This plasmid comprises the barnase gene, which acts as  
 a male-sterility gene. The plasmid is used to create transgenic plants  
 of the invention.

Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atagggtggaggctatttgg 21

|||||

Db 4716 ATAGGTTGGAGGCTATTGG 4696

RESULT 6

AAT33339/c

ID AAT33339 standard; DNA; 5864 BP.

XX AC AAT33339;

XX DT 22-JAN-1997 (first entry)

XX Plasmid pTCO13 T-DNA used to obtain male sterile oilseed rape.

XX Plasmid pTCO13; male sterile; barnase; ribonuclease; barstar;

XX transgenic plant; oilseed rape; canole; Brassica napus; ds.

OS Synthetic.

XX Key

FH Location/Qualifiers



```

FT FT misc_feature /note= "Synthetic polylinker derived sequence"
FT FT 2660..2920
FT FT /*tag= i
FT FT /note= "Tag1 fragment from the 3' untranslated end of the
FT FT nopaline synthase gene (3'nos) from the T-DNA of pTir37
FT FT and containing plant polyadenylation signals"
FT FT 2921..2936
FT FT /*tag= j
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 2937..3032
FT FT /*tag= k
FT FT /note= "The 3' untranslated region downstream from the
FT FT barnase coding sequence of Bacillus amyloliquefaciens"
FT FT 3033..3368
FT FT /*tag= l
FT FT /product= "Protein encoded by barnase gene from
FT FT Bacillus amyloliquefaciens"
FT FT 3369..4878
FT FT /*tag= m
FT FT /note= "Promoter region of the anther-specific gene
FT FT TA29 from Nicotiana tabacum"
FT FT 4879..4924
FT FT /*tag= n
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 4925..5215
FT FT /*tag= o
FT FT /note= "Promoter of the nopaline synthase gene from the
FT FT T-DNA of pTir37 of Agrobacterium tumefaciens"
FT FT 5216..5217
FT FT /*tag= p
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5218..5490
FT FT /*tag= q
FT FT /product= "Protein encoded by barstar gene of
FT FT Bacillus amyloliquefaciens"
FT FT 5491..5530
FT FT /*tag= r
FT FT /note= "Sequence from the 3' untranslated end of the
FT FT barstar gene from Bacillus amyloliquefaciens"
FT FT 5531..5554
FT FT /*tag= s
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5555..5766
FT FT /*tag= t
FT FT /note= "The 3' untranslated end from the TL-DNA
FT FT gene 7 (3'g7) of pTiB6S3"
FT FT 5767..5773
FT FT /*tag= u
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5774..5810
FT FT /*tag= v
FT FT /note= "Residual sequence from the TL-DNA at the
FT FT right border repeat"
FT FT 5811..5840
FT FT /*tag= w
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5841..5865
FT FT /*tag= x
FT FT /note= "Left border repeat from the TL-DNA from
FT FT pTiB6S3"
FT FT
XX PN WO200131042-A2.
XX XX
XX PD 03-MAY-2001.
XX XX
XX PF 26-OCT-2000; 2000WO-EP10680.
XX XX
XX PR 29-OCT-1999; 99US-0430497.
XX XX
XX PA (AVET ) AVENTIS CROPS SCIENCE NV.
XX XX
XX PI Weston B, De Beuckeleer M;
XX XX

```

```

DR WPI; 2001-300517/31.
XX
FT Transgenic Brassica plants, seeds, cells or tissues, characterized by
FT harboring specific transformation events, particularly by presence of
FT male-sterility gene, at specific location in its genome -
XX
PS Claim 1; Page 47-49; 53pp; English.
XX
CC The present invention relates to a transgenic Brassica plant or its
CC seed, cells or tissues, characterised by harbouring a specific
CC transformation event, particularly by the presence of a male-sterility
CC gene, at a specific location in the Brassica genome. Transgenic
CC Brassica plant is useful for producing a hybrid seed by crossing the
CC transgenic plant with a male-fertile Brassica plant and harvesting the
CC hybrid seed from the transgenic Brassica plant.
CC The present sequence is chimeric T-DNA of plasmid pTir37. This sequence
CC comprises right border repeat, left border repeat and 3' untranslated
CC region (UTR) from TL-DNA of pTiB6S3, synthetic polylinker sequences,
CC coding regions of blalaphos resistance gene (bar) from
CC Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens
CC and barstar gene from Bacillus amyloliquefaciens and promoters of atS1A
CC ribulose-1,5-bisphosphate carboxylase small subunit gene from
CC Arabidopsis thaliana, the anther-specific gene TA29 from
CC Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTir37
CC of Agrobacterium tumefaciens.
XX
SQ Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;

Query Match 100.0%; Score 21; DB 22; Length 5865;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
Db 4717 ATAGGTGGGAGGCTATTGG 4697
|||||
RESULT 8
AAF83820/c
ID AAF83820 standard; DNA; 1983 BP.
XX
AC AAF83820;
XX
DT 23-JUL-2001 (first entry)
DE Dengue-2 virus derived DNA fragment.
XX
KW Yellow fever virus; prM-E protein; flavivirus; chimeric; medicament;
KW infection; tumor antigen; cytokine; lymphoid; reticuloendothelial;
KW cancer; virucide; vaccine; dengue virus; ds.
XX
OS Dengue virus.
XX
FH Key Location/Qualifiers
FT CDS 1..1983
FT /*tag= a
XX
XX WO200139802-A1.
XX
XX PD 07-JUN-2001.
XX
XX PF 01-DEC-2000; 2000WO-US32821.
XX
XX PR 01-DEC-1999; 99US-0452638.
XX
XX PA (ORAV-) ORAVAX INC.
XX
XX PI Chambers TJ, Monath TP, Guirakhoo F;
XX
XX DR WPI; 2001-343953/36.
XX
XX DR P-PSDB; AAB84901.
XX

```

FT Chimeric live, infectious, attenuated yellow fever viruses used for  
 PT preventing and treating diseases caused by flaviviruses have prM-E  
 PT nucleotide sequence from a second, different flavivirus as functional  
 PT yellow fever prM-E is not expressed -  
 XX  
 PS Disclosure; Page 196-198; 232pp; English.  
 XX  
 CC The invention relates to a chimeric live, infectious, attenuated virus  
 CC comprising a yellow fever virus with the nucleotide sequence encoding a  
 CC prM-E protein deleted, truncated or mutated so that functional yellow  
 CC fever virus prM-E protein is not expressed and also integrated into the  
 CC genome of the yellow fever virus a nucleotide sequence encoding a prM-E  
 CC protein of a second, different flavivirus so that the prM-E protein of  
 CC the second flavivirus is expressed. The chimeric live, infectious,  
 CC attenuated virus is used to prepare medicaments for preventing or  
 CC treating flavivirus infection in a patient. The yellow fever virus vector  
 CC produces its gene product (tumor antigen or cytokine) in cells of the  
 CC lymphoid or reticuloendothelial system or in a precursor of these systems  
 CC in patients with cancer. Flaviviruses replicate in the cytoplasm of cells  
 CC so that the virus replication does not involve integration of the viral  
 CC genome into the host cell. The present sequence represents a DNA fragment  
 CC derived from dengue-2 virus.  
 XX  
 SQ Sequence 1983 BP; 645 A; 409 C; 509 G; 420 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 1983;  
 Best Local Similarity 90.0%; Pred. No. 42;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 taggggtgggaggtatttgg 21  
 ||||| ||||| ||||| |||||  
 Db 666 TAGGGTGGCAGCGCTGTTGG 647

RESULT 9  
 AAT49304/C  
 ID AAT49304 standard; cDNA; 10723 BP.  
 XX  
 AC AAT49304;  
 XX  
 DT 12-SEP-1997 (first entry)  
 XX  
 DE cDNA encoding polyprotein of attenuated DEN-2 virus, PDK-53.  
 KW Dengue 2 virus; polyprotein; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;  
 KW NS4A; NS4B; NS5; PDK-53; quadravalent vaccine; immunity; serotype;  
 KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;  
 KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;  
 KW DHF; DSS; ss.  
 XX  
 OS Dengue 2 virus, strain 16681.  
 OS Synthetic.  
 XX  
 FH Key . Location/Qualifiers  
 FT CDS 97..10272  
 FT /\*tag= a  
 FT /product= DEN-2 attenuated polyprotein  
 FT /transl\_except(pos:643..645,aa:Xaa)  
 FT /transl\_except(pos:1135..1137,aa:Xaa)  
 FT /transl\_except(pos:1393..1395,aa:Xaa)  
 FT /transl\_except(pos:2809..2811,aa:Xaa)  
 FT /transl\_except(pos:3040..3042,aa:Xaa)  
 FT /transl\_except(pos:9208..9210,aa:Xaa)  
 FT /note= "Xaa = unknown amino acid"  
 FT 57  
 FT mutation  
 FT /\*tag= b  
 FT /note= "C>T mutation"  
 FT 524  
 FT mutation  
 FT /\*tag= c  
 FT /note= "A>T mutation, causes Asp to Val substitution"  
 FT 2055  
 FT mutation  
 FT /\*tag= d

FT mutation  
 FT /note= "C>T mutation"  
 FT 2579  
 FT /\*tag= e  
 FT /note= "G>A mutation, causes Gly to Asp substitution"  
 FT 4018  
 FT /\*tag= f  
 FT /note= "C>T mutation, causes Leu to Phe substitution"  
 FT 5547  
 FT /\*tag= g  
 FT /note= "C>T mutation"  
 FT 6599  
 FT /\*tag= h  
 FT /note= "G>C mutation, causes Gly to Ala substitution"  
 FT 8571  
 FT /\*tag= i  
 FT /note= "C>T mutation"  
 XX  
 XX WO9640933-A1.  
 XX  
 XX 19-DEC-1996.  
 PD  
 XX  
 XX 06-JUN-1996; 96WO-US09209.  
 PF  
 XX  
 XX 07-JUN-1995; 95US-0483292.  
 PR  
 XX  
 PA (UYMA-) UNIV MAHIDOL AT SALAYA.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 XX Bhamarapravati N. Butrapet S, Chang J, Gubler DJ;  
 PI Halstead SB, Kinney R, Trent DW;  
 XX  
 DR WPI; 1997-052330/05.  
 DR P-PSDB; AAW06591.  
 XX  
 XX PDK-53, a clone of infectious attenuated Dengue 2 virus strain 16681  
 PT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a  
 PT quadravalent vaccine for protecting against Dengue virus infection  
 XX  
 PS Claim 27; Page 122-136; 261pp; English.  
 XX  
 CC This sequence encodes the polyprotein from an attenuated derivative  
 CC of Dengue 2 virus, strain 16681. The derivative is designated PDK-53.  
 CC The polyprotein comprises the capsid, prM, M, E, NS1, NS2A, NS2B, NS3,  
 CC NS4A, NS4B and NS5 proteins. The PDK-53 viral sequence may be used in  
 CC the production of a quadravalent vaccine which provides immunity against  
 CC all four serotypes of dengue virus. The vaccine also comprises a  
 CC chimeric DEN-2/1 virus, a chimeric DEN-2/3 virus, and/or a chimeric  
 CC DEN-2/4 virus. The new quadravalent vaccines are used to protect  
 CC against infection by all four serotypes of dengue virus, DEN-1, DEN-2,  
 CC DEN-3 and DEN-4, which can lead to dengue fever or fatal dengue  
 CC haemorrhagic fever/dengue shock syndrome (DHF/DSS). Host cells are  
 CC used to produce the recombinant protein products of the DNA constructs  
 CC which are used in the vaccines.  
 XX  
 SQ Sequence 10723 BP; 3548 A; 2196 C; 2711 G; 2261 T; 7 other;

Query Match 80.0%; Score 16.8; DB 18; Length 10723;  
 Best Local Similarity 90.0%; Pred. No. 48;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 taggggtgggaggtatttgg 21  
 ||||| ||||| ||||| |||||  
 Db 1104 TAGGGTGGCAGCGCTGTTGG 1085  
 RESULT 10  
 AAT49303/C  
 ID AAT49303 standard; cDNA; 10723 BP.  
 XX  
 AC AAT49303;  
 XX  
 DT 11-SEP-1997 (first entry)

```

XX DE cDNA sequence encoding polyprotein of DEN-2 virus, strain 16681.
XX KW Dengue 2 virus; polyprotein; capsid; prM; M; E; NS1; NS2A; NS2B; NS3;
XX NS4A; NS4B; NS5; pDK-53; quadravalent vaccine; immunity; serotype;
XX KW chimeric DEN-2/1 virus; chimeric DEN-2/3 virus; chimeric DEN-2/4 virus;
XX KW dengue fever; fatal dengue haemorrhagic fever; dengue shock syndrome;
XX DHF; DSS; ss.
XX OS Dengue 2 virus, strain 16681.
XX FH Key
XX FT Location/Qualifiers
XX CDS 97..10272
XX FT /*tag= a
XX FT /product= DEN-2 polyprotein
XX FT /transl_except(pos:9208..9210, aa:Xaa)
XX FT /note= "Xaa = unknown amino acid"
XX PN WO9640933-A1.
XX PD 19-DEC-1996.
XX PP 06-JUN-1996; 96WO-US09209.
XX PF 07-JUN-1995; 95US-0483292.
XX PR (UYMA-) UNIV MAHIDOL AT SALAYA.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Bhamarapravati N, Butrapet S, Chang J, Gubler DJ;
XX PI Halstead SB, Kinney R, Trent DW;
XX DR WPT; 1997-052330/05.
XX DR P-PSDB; AAW06590.
XX XX
XX PT PDK-53, a clone of infectious attenuated Dengue 2 virus strain 16681
XX PT - also chimeric DEN-2/1, DEN-2/3 and DEN-1/4 viruses, used as a
XX PT quadravalent vaccine for protecting against Dengue virus infection
XX PS Claim 23; Page 107-121; 261pp; English.
XX CC This sequence encodes the polyprotein from Dengue 2 virus, strain 16681.
XX CC The polyprotein comprises the capsid, prM, M, E, NS1, NS2A, NS2B, NS3,
XX CC NS4A, NS4B and NS5 proteins. A clone of this wildtype viral sequence,
XX CC PDK-53, may be used in the production of a quadravalent vaccine which
XX CC provides immunity against all four serotypes of dengue virus. The
XX CC vaccine also comprises a chimeric DEN-2/1 virus, a chimeric DEN-2/3
XX CC virus, and/or a chimeric DEN-2/4 virus. The new quadravalent vaccines
XX CC are used to protect against infection by all four serotypes of dengue
XX CC virus, DEN-1, DEN-2, DEN-3 and DEN-4, which can lead to dengue fever or
XX CC fatal dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS). Host
XX CC cells are used to produce the recombinant protein products of the DNA
XX CC constructs which are used in the vaccines.
XX SQ Sequence 10723 BP; 3553 A; 2200 C; 2713 G; 2255 T; 2 other;

Query Match 80.0%; Score 16.8; DB 18; Length 10723;
Best Local Similarity 90.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggggtggaggcgtattgg 21
||||| |||||
DB 1104 TAGGGTGGCAGGCTGTTGG 1085

RESULT 11
AAF68287
ID AAF68287 standard; cDNA; 694 BP.
XX
AC AAF68287;
XX
DT 12-APR-2001 (first entry)

```

```

XX DE Human lung tumour protein related nucleotide sequence SEQ ID NO:205.
XX KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
XX KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
XX KW cytostatic; antisense inhibition; ss.
XX XX
XX OS Homo sapiens.
XX PN WO200100828-A2.
XX PD 04-JAN-2001.
XX PF 30-JUN-2000; 2000WO-US18061.
XX PR 30-JUN-1999; 99US-0346492.
XX PR 15-OCT-1999; 99US-0419356.
XX PR 17-DEC-1999; 99US-0466867.
XX PR 30-DEC-1999; 99US-0476300.
XX PR 06-MAR-2000; 2000US-0519642.
XX PR 22-MAR-2000; 2000US-0533077.
XX PR 10-APR-2000; 2000US-0546259.
XX PR 27-APR-2000; 2000US-0560406.
XX PR 05-JUN-2000; 2000US-0589184.
XX PA (CORI-) CORIXA CORP.
XX XX
XX PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX PI Retter MW, Mannion J;
XX XX
XX DR WPI; 2001-071488/08.
XX XX
XX PT Lung tumor-associated proteins and the nucleic acids that encode them,
XX PT useful for preventing, diagnosing and treating lung cancer -
XX PS Claim 4; Page 213; 436pp; English.
XX CC The present invention describes immunogenic portions of lung tumour-
XX CC associated proteins (I) and the nucleic acids (NAs) that encode them.
XX CC (I) have cytostatic activity and can be used in gene therapy, antisense
XX CC inhibition and in vaccines. The NAs and the lung tumour-associated
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with their inappropriate expression,
XX CC especially lung cancers. For example, the NAs may be administered to
XX CC treat diseases by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of the protein by expressing inactive proteins
XX CC or to supplement the patient's own production of (I). Additionally, the
XX CC NAs may be used to produce the lung-tumour associated protein, according
XX CC to standard recombinant DNA methodology. Conversely, antisense NA
XX CC molecules may be administered to down regulate protein expression by
XX CC binding with the cells own genes and preventing their expression. The NA
XX CC and complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar NA sequences in
XX CC samples, and hence which patients may be in need of treatment for lung
XX CC cancer. The (I) may be used as antigens in the production of antibodies
XX CC and in assays to identify modulators (agonists and antagonists) of the
XX CC expression and activity of the protein. AAF68083 to AAF68878 and
XX CC AAB76848 to AAB76878 represent human lung tumour protein related
XX CC nucleotide and protein sequences which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 694 BP; 202 A; 111 C; 154 G; 227 T; 0 other;

```

```

Query Match 78.1%; Score 16.4; DB 22; Length 694;
Best Local Similarity 94.4%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 aggggtggaggcgtattgg 20
||||| ||||| |||
DB 290 aggggtggaggcgtattgg 307

```

```
RESULT 12
AAQ46850/c
ID AAQ46850 standard; cDNA; 857 BP.
XX AC AAQ46850;
XX AC
XX DT 26-JAN-1994 (first entry)
XX DE Recombinant human kappa casein gene.
XX KW Casein; supplement; milk; pharmaceutical; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 45..593
XX FT /*tag= a
XX FT
XX PN WO9608269-A1.
XX PD 21-MAR-1996.
XX XX
XX XX 05-MAY-1995; 95WO-US05676.
XX PF 16-SEP-1994; 94US-0308883.
XX PR 16-SEP-1994; 94US-0308882.
XX XX
XX PA (ABBO ) ABBOTT LAB.
XX PI Baxter JH, Cummings RD, Mukerji P, Prieto PA, Seo AE;
XX WPI; 1996-179724/18.
XX DR
XX XX
XX PT Kappa-casein used as anti-rotaviral infection agent in nutritional
XX PT product - to prevent, retard or treat rotavirus infection,
XX PT especially in infants, children and immuno-deficient patients
XX PS Claim 1; Fig 8; 42pp; English.
XX CC
XX CC AAT14564 encodes human milk kappa-casein which is useful as a component
XX CC of an enteral nutritional product. Kappa-casein has anti-rotaviral
XX CC activity, it inhibits the attachment of human rotavirus to mammalian
XX CC cells. The kappa-casein is useful in the treatment and prevention of
XX CC onset of gastroenteritis and other diarrhoeal diseases caused by
XX CC rotaviridae esp. in infants, children or immunodeficient patients.
XX CC Kappa-casein appeared to have no side effects on the subjects treated
XX CC and is unlikely to be toxic or cause an allergic reaction.
XX SQ Sequence 857 BP; 309 A; 219 C; 127 G; 202 T; 0 other;

Query Match 78.1%; Score 16.4; DB 17; Length 857;
Best Local Similarity 94.4%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 gggtggggaggctattgg 21
||||| |||||||
Db 359 GGGTGGGTGGCTATTGG 342

RESULT 14
AAQ89598/c
ID AAQ89598 standard; cDNA; 858 BP.
XX AC AAQ89598;
XX XX
XX DT 06-NOV-1995 (first entry)
XX DE Human kappa-casein cDNA.
XX KW Kappa-casein; milk protein; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 107..655
XX FT /*tag= a
XX PN US5391497-A.
XX PD 21-FEB-1995.

Query Match 78.1%; Score 16.4; DB 14; Length 857;
Best Local Similarity 94.4%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 gggtggggaggctattgg 21
||||| |||||||
Db 359 GGGTGGGTGGCTATTGG 342

RESULT 13
AAT14564/c
ID AAT14564 standard; cDNA; 857 BP.
XX AC AAT14564;
XX XX
XX DT 24-OCT-1996 (first entry)
XX DE Human milk kappa-casein cDNA.
```

XX 13-OCT-1992; 92US-0962569.  
XX 13-OCT-1992; 92US-0962569.  
XX (COLS ) UNIV COLORADO FOUND INC.  
XX Chang Y, Ham RG, Jeffers KF, Menon RS;  
XX WPI; 1995-160470/21.  
XX P-PSDB; AAR72696.  
XX DNA encoding human kappa-casein - used for the prodn. of large amts. of  
XX highly purified kappa-casein milk protein for infant use.  
XX Claim 4; Column 13-16; 14pp; English.  
XX A commercial cDNA library prepd. in lambda gtil from mRNA obtd. from  
XX human breast tissue removed during the third trimester of pregnancy was  
XX screened with rabbit anti-bovine kappa-casein cDNA. The cDNA  
XX insert of a recombinant phage was amplified by PCR. The sequence  
XX of an isolated full-length clone encoding human kappa-casein  
XX (AAR72696) is given in AAQ89598.  
XX Sequence 858 BP; 266 A; 237 C; 137 G; 218 T; 0 other;  
SQ  
Query Match 78.1%; Score 16.4; DB 16; Length 858;  
Best Local Similarity 94.4%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 ggggtggaggctattgg 21  
DB 421 GGGTGGGTGGCTATTGG 404  
RESULT 15  
AAZ21101/c  
ID AAZ21101 standard; cDNA; 2185 BP.  
AC AAZ21101;  
DT 17-NOV-1999 (first entry)  
DE Human secreted protein clone qb401\_6 encoding cDNA.  
KW Human; secreted protein; biological activity; nutritional; cytokine;  
KW cell proliferation; differentiation; immune stimulating; vaccine;  
KW hematopoiesis regulation; tissue growth; haemostatic; thrombolytic;  
KW anti-inflammatory; tumour inhibition; ss.  
OS Homo sapiens.  
XX WO9946287-A1.  
XX 16-SEP-1999.  
XX 11-MAR-1999; 99WO-US05243.  
XX 11-MAR-1998; 98US-0077521.  
XX 14-MAY-1998; 98US-0079124.  
XX 10-MAR-1999; 99US-0266105.  
XX (CEMY ) GENETICS INST INC.  
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ;  
XX WPI; 1999-551362/46.  
XX P-PSDB; AAY29869.  
XX Polynucleotides encoding secreted human proteins, derived from human  
XX fetal brain, human adult blood, human adult bladder, or human adult

PT neural tissue cDNA libraries.  
XX Claim 26; Page 109; 118pp; English.  
XX AAZ21093 to AAZ21102 encode new human secreted proteins and AAY29861 to  
XX AAY29873 represent the secreted proteins encoded by the polynucleotide  
XX sequences. AAZ21103 to AAZ21112 represent probes for the secreted  
XX proteins. The polynucleotides and proteins are predicted to have  
XX biological activities which would make them suitable for treating,  
XX preventing or ameliorating medical conditions in humans and animals,  
XX although no supporting data is given. Suggested activities include  
XX nutritional activity, cytokine and cell proliferation/differentiation  
XX activity, immune stimulating activity, tissue growth activity,  
XX hematopoiesis regulating activity, chemotactic/chemokinetic activity, haemostatic  
XX activin/inhibin activity, receptor/ligand activity, anti-inflammatory  
XX activity, cadherin/tumour invasion suppressor activity, and tumour  
XX inhibition activity. The polynucleotides and proteins can also be used  
XX as nutritional sources or supplements. Such uses include use as a protein  
XX or amino acid supplement, use as a carbon source, use as a nitrogen  
XX source and use as a source of carbohydrate. They may also have utility  
XX in compositions used for bone, cartilage, tendon, ligament, and/or nerve  
XX tissue growth or regeneration, as well as for wound healing and tissue  
XX repair and replacement, and in the treatment of burns, incisions and  
XX ulcers. The proteins which induce cartilage and/or bone growth in  
XX circumstances where bone is not normally formed, have application in  
XX the healing of bone fractures and cartilage damage or defects in humans  
XX and other animals.  
XX Sequence 2185 BP; 472 A; 572 C; 548 G; 593 T; 0 other;  
SQ  
Query Match 78.1%; Score 16.4; DB 20; Length 2185;  
Best Local Similarity 94.4%; Pred. No. 66;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 taggggtggaggctattt 19  
DB 1052 TAGGGTGGAGGCTATT 1035

Search completed: February 25, 2002, 18:17:21  
Job time: 16679 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:08 ; Search time 301.6 Seconds  
(without alignments)  
15.769 Million cell updates/sec

Title: US-09-698-903B-3

Perfect score: 21

Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21	100.0	4946	US-08-817-188-1	Sequence 1, Appli
C 2	21	100.0	5864	US-08-894-440-4	Sequence 4, Appli
C 3	16.4	78.1	857	US-08-308-883-1	Sequence 1, Appli
C 4	16.4	78.1	857	US-08-730-163-1	Sequence 1, Appli
C 5	16.4	78.1	857	US-08-256-799-1	Sequence 1, Appli
C 6	16.4	78.1	857	US-08-462-437-1	Sequence 1, Appli
C 7	16.4	78.1	858	US-07-962-569A-7	Sequence 7, Appli
C 8	16.4	78.1	9844	US-08-462-437-30	Sequence 30, Appli
C 9	16.4	78.1	13104	US-08-256-799-4	Sequence 4, Appli
C 10	16.4	78.1	13104	US-08-462-437-4	Sequence 4, Appli
C 11	16.2	77.1	1418	US-09-111-470-7	Sequence 7, Appli
C 12	16.2	77.1	24979	US-08-147-777-3	Sequence 3, Appli
C 13	16.2	77.1	24979	US-08-452-872-3	Sequence 3, Appli
C 14	16.2	77.1	24979	PCT-US93-03985-3	Sequence 3, Appli
C 15	15.8	75.2	1025	US-08-136-277-23	Sequence 23, Appli
C 16	15.8	75.2	1025	US-08-479-403-23	Sequence 23, Appli
C 17	15.8	75.2	1025	US-08-835-734-23	Sequence 23, Appli
C 18	15.8	75.2	1643	US-08-805-118-2	Sequence 2, Appli
C 19	15.8	75.2	1780	US-08-724-394A-19	Sequence 19, Appli
C 20	15.4	73.3	5769	US-08-652-971-1	Sequence 1, Appli
C 21	15.4	73.3	5769	US-08-991-258A-1	Sequence 1, Appli
C 22	15.4	73.3	5769	US-08-769-399-1	Sequence 1, Appli
C 23	15.4	73.3	5769	US-08-991-953A-1	Sequence 1, Appli
C 24	15.2	72.4	5538	US-08-231-193A-55	Sequence 55, Appli
C 25	15.2	72.4	5538	US-08-486-273A-55	Sequence 55, Appli
C 26	15.2	72.4	5538	US-08-940-086A-55	Sequence 55, Appli
C 27	15.2	72.4	9046	US-08-227-536-1	Sequence 1, Appli

C 28	15.2	72.4	9046	5	PCT-US95-04682-1	Sequence 1, Appli
C 29	14.8	70.5	354	2	US-08-676-279-38	Sequence 38, Appli
C 30	14.8	70.5	569	4	US-09-326-039-13	Sequence 13, Appli
C 31	14.8	70.5	2750	1	US-08-136-277-1	Sequence 1, Appli
C 32	14.8	70.5	2750	2	US-08-479-403-1	Sequence 1, Appli
C 33	14.8	70.5	2750	3	US-08-835-734-1	Sequence 1, Appli
C 34	14.6	69.5	35100	1	US-08-306-691B-19	Sequence 19, Appli
C 35	14.6	69.5	35100	5	PCT-US93-06251-19	Sequence 19, Appli
C 36	14.4	68.6	90050	4	US-09-245-041-5	Sequence 5, Appli
C 37	14.2	67.6	2899	2	US-08-624-581-2	Sequence 2, Appli
C 38	14.2	67.6	2959	2	US-08-624-581-1	Sequence 1, Appli
C 39	14.2	67.6	5994	3	US-09-032-365A-11	Sequence 11, Appli
C 40	14.2	67.6	13146	2	US-08-724-354D-3	Sequence 3, Appli
C 41	14.2	67.6	13146	3	US-09-270-984A-3	Sequence 3, Appli
C 42	14.2	67.6	14311	4	US-08-646-695-1	Sequence 1, Appli
C 43	14.2	67.6	14311	4	US-08-646-695-7	Sequence 7, Appli
C 44	14.2	67.6	14311	5	PCT-US96-06053-1	Sequence 1, Appli
C 45	14.2	67.6	14311	5	PCT-US96-06053-7	Sequence 7, Appli

## ALIGNMENTS

RESULT 1

US-08-817-188-1/c  
; Sequence 1, Application US/08817188  
; Patent No. 6074876  
; GENERAL INFORMATION:  
; APPLICANT: DE BLOCK, MARC  
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
; FILE REFERENCE: 2121-0127P  
; CURRENT APPLICATION NUMBER: US/08/817,188  
; CURRENT FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
; EARLIER FILING DATE: 1996-07-31  
; EARLIER APPLICATION NUMBER: EP 95401844.6  
; EARLIER FILING DATE: 1995-08-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 4946  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pTHW107

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((1)..(25))  
OTHER INFORMATION: T-DNA right border (RB)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((97)..(330))  
OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((331)..(882))  
OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
OTHER INFORMATION: transferase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((883)..(2608))  
OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((2658)..(3031))  
OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
OTHER INFORMATION: gene of Agrobacterium T-DNA  
FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4876))
; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((4922)..(4946))
; OTHER INFORMATION: LB: T-DNA left border
US-08-817-188-1

Query Match      100.0%; Score 21; DB 3; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.056;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
   |||||||||||||||||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 2
US-08-894-440-4/c
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pICOLL3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens

```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Phos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

Query Match      100.0%; Score 21; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 21
   |||||||||||||||||||
Db 4716 ATAGGTGGGAGGCTATTGG 4696

RESULT 3
US-08-308-883-1/c
; Sequence 1, Application US/08308883
; Patent No. 5576300
; GENERAL INFORMATION:
; APPLICANT: Mukerji, P. A.
; APPLICANT: Prieto, P. A.
; APPLICANT: Seo, A. E.-Y.
; APPLICANT: Baxter, J. H.
; APPLICANT: Cummings, R.D.
; TITLE OF INVENTION: Method for Inhibition of Human Rotavirus Infection.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lonnie R. Drayer
; ADDRESSEE: ROSS Products Division
; ADDRESSEE: Abbott Laboratories
; STREET: 625 Cleveland Avenue
; CITY: Columbus
; STATE: Ohio
; COUNTRY: United States
; ZIP: 43215
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh System 7.1
; SOFTWARE: ClarisWorks 1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,883
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: No. 5576300 applicable
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (614) 624-3774
; TELEFAX: (614) 624-3074
; TELEX: No. 5576300e
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```
; LENGTH: 857 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human milk kappa-casein
; HYPOTHETICAL: NO
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: Human
; ORGANISM: Homo sapiens
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: Adult
; HAPLOTYPE:
; TISSUE TYPE: Mammary gland
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE: Human Mammary Gland
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 45...593
; IDENTIFICATION METHOD: DNA sequencing and restriction analysis
; OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the human
; PUBLIC INFORMATION:
; AUTHORS: L. Hansson et al
; TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use The
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER: PCT/WO93/15196
; FILING DATE: 25-JAN-1993
; PUBLICATION DATE: 05-AUG-1993
; RELEVANT RESIDUES IN SEQ ID NO:
;
; US-08-308-883-1

Query Match 78.1%; Score 16.4; DB 1; Length 857;
Best Local Similarity 94.4%; Pred. No. 9.5;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggggtggaggctattgg 21
||||| |||||||
Db 359 GGGTGGGTGGCTATTGG 342

RESULT 4
US-08-730-163-1/c
; Sequence 1, Application US/08730163
; Patent No. 5712250
; GENERAL INFORMATION:
; APPLICANT: Mukerji, P.
; APPLICANT: Prieto, P. A.
; APPLICANT: Seo, A. E.-Y.
; APPLICANT: Baxter, J. H.
; APPLICANT: Cummings, R.D.
; TITLE OF INVENTION: Product for Inhibition of Human Rotavirus Infection.
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lonnie R. Drayer
; ADDRESSEE: ROSS Products Division
; ADDRESSEE: Abbott Laboratories
; STREET: 625 Cleveland Avenue
; CITY: Columbus
```

```
; STATE: Ohio
; COUNTRY: United States
; ZIP: 43215
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb storage (B)COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh System 7.1(D)SOFTWARE: ClarisWorks 1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/730,163
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,882
; FILING DATE: 16-SEP-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (614) 624-3774
; TELEFAX: (614) 624-3074
; TELEX: No. 5712250e
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human milk kappa-casein
; HYPOTHETICAL: NO
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: Human
; ORGANISM: Homo sapiens
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: Adult
; HAPLOTYPE:
; TISSUE TYPE: Mammary gland
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE: Human Mammary Gland
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 45...593
; IDENTIFICATION METHOD: DNA sequencing and restriction analysis
; OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the hu
; PUBLIC INFORMATION:
; AUTHORS: L. Hansson et al
; TITLE: DNA Encoding Kappa-Casein, Process for Obtaining the Protein and Use
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER: PCT/WO93/15196
; FILING DATE: 25-JAN-1993
; PUBLICATION DATE: 05-AUG-1993
; RELEVANT RESIDUES IN SEQ ID NO:
;
; US-08-730-163-1

Query Match 78.1%; Score 16.4; DB 1; Length 857;
Best Local Similarity 94.4%; Pred. No. 9.5;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggggtggaggctattgg 21
||||| |||||||
Db 359 GGGTGGGTGGCTATTGG 342
```

RESULT 5  
US-08-256-799-1/C  
; Sequence 1, Application US/08256799  
; Patent No. 6222094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROEM, Sven  
; APPLICANT: HERNELL, Olie  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/256,799  
; FILING DATE: 06-DEC-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON=1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 45..104  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 13..44  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: 594..848  
US-08-256-799-1

Query Match 78.1%; Score 16.4; DB 4; Length 857;  
Best Local Similarity 94.4%; Pred. No. 9.5;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gggtgggaggtattgg 21

Db 359 GGTTGGTGGCTATTGG 342  
||||| |||||||  
RESULT 6  
US-08-462-437-1/C  
; Sequence 1, Application US/08462437  
; Patent No. 6232094  
; GENERAL INFORMATION:  
; APPLICANT: HANSSON, Lennart  
; APPLICANT: STROEMQVIST, Mats  
; APPLICANT: BERGSTROEM, Sven  
; APPLICANT: HERNELL, Olie  
; APPLICANT: TOERNELL, Jan  
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/462,437  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DK 88/92  
; FILING DATE: 23-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: HANSSON=1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 857 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 45..593  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 45..104  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 13..44  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: 594..848  
US-08-462-437-1

Query Match 78.1%; Score 16.4; DB 4; Length 857;  
Best Local Similarity 94.4%; Pred. No. 9.5;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggggtggaggctattgg 21  
 ||||| |||||  
 Db 359 GGGTGGTGGCTATTGG 342

## RESULT 7

US-07-962-569A-7/c  
 ; Sequence 7, Application US/07962569A  
 ; Patent No. 5391497  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MENON, RAVI S.  
 ; APPLICANT: JEFFERS, KATHLEEN F.  
 ; APPLICANT: CHANG, YING-FON  
 ; APPLICANT: HAM, RICHARD G.  
 ; TITLE OF INVENTION: HUMAN K-CASEIN  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FREDERICK W. PEPPER, PH.D.  
 ; STREET: 11545 W. BERNARDO COURT, STE. 302  
 ; CITY: SAN DIEGO  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92127

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/962.569A  
 FILING DATE: 19921013  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
 NAME: PEPPER PH.D., FREDERICK W.  
 REGISTRATION NUMBER: 31,286  
 REFERENCE/DOCKET NUMBER: 920224.01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 451-1120  
 TELEFAX: (619) 451-9628  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 858 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 107..656  
 US-07-962-569A-7

Query Match 78.1%; Score 16.4; DB 1; Length 858;  
 Best Local Similarity 94.4%; Pred. No. 9.5;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggggtggaggctattgg 21  
 ||||| |||||  
 Db 421 GGGTGGTGGCTATTGG 404

## RESULT 8

US-08-462-437-30/c  
 ; Sequence 30, Application US/08462437  
 ; Patent No. 6232094  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HANSSON, Lennart  
 ; APPLICANT: STROMQVIST, Mats  
 ; APPLICANT: BERGSTROM, Sven  
 ; APPLICANT: HERNELL, Olle  
 ; APPLICANT: TOERNELL, Jan  
 ; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS

; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF  
 ; NUMBER OF SEQUENCES: 31  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BROWDY AND NEIMARK  
 ; STREET: 419 Seventh Street, N.W., Suite 300  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20004  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/462.437  
 ; FILING DATE: 05-JUN-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: DK 88/92  
 ; FILING DATE: 23-JAN-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: COOPER, Iver P.  
 ; REGISTRATION NUMBER: 28,005  
 ; REFERENCE/DOCKET NUMBER: HANSSON-1A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-628-5197  
 ; TELEFAX: 202-737-3528  
 ; TELEX: 248633  
 ; INFORMATION FOR SEQ ID NO: 30:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9844 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 362..425  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 426..2571  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 2572..2633  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 2634..5575  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 5576..5608  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 5609..5754  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 5755..7250  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 7251..9017  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: 9018..9184  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: join(2580..2633, 5576..5608, 6755..7216)  
 ; US-08-462-437-30

Query Match 78.1%; Score 16.4; DB 4; Length 9844;

```
Best Local Similarity 94.4%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggggtggaggctatttgg 21
||||| |||||||
Db 6982 GGGTGGGCTATTGG 6965

RESULT 9
US-08-256-799-4/c
; Sequence 4, Application US/08256799
; Patent No. 6222094
; GENERAL INFORMATION:
; APPLICANT: HANSSON, Lennart
; APPLICANT: STROEMQVIST, Mats
; APPLICANT: BERGSTROEM, Sven
; APPLICANT: HERNELL, Olle
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS FOR
; TITLE OF INVENTION: OBTAINING THE PROTEIN AND USE THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 06-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 88/92
; FILING DATE: 23-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: HANSSON-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: p9459
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..8834
; FEATURE:
; NAME/KEY: intron
; LOCATION: 8868..10014
; FEATURE:
; NAME/KEY: intron
; LOCATION: 10511..12277
; FEATURE:
; NAME/KEY: exon
; LOCATION: 8835..8867
; FEATURE:
; NAME/KEY: exon
; LOCATION: 10015..10510

Query Match 78.1%; Score 16.4; DB 4; Length 13104;
Best Local Similarity 94.4%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ggggtggaggctatttgg 21
||||| |||||||
Db 10242 GGGTGGGCTATTGG 10225

RESULT 10
US-08-462-437-4/c
; Sequence 4, Application US/08462437
; Patent No. 6232094
; GENERAL INFORMATION:
; APPLICANT: HANSSON, Lennart
; APPLICANT: STROEMQVIST, Mats
; APPLICANT: BERGSTROEM, Sven
; APPLICANT: HERNELL, Olle
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: DNA ENCODING KAPPA-CASEIN, PROCESS
; TITLE OF INVENTION: FOR OBTAINING THE PROTEIN AND USE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,437
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 88/92
; FILING DATE: 23-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: HANSSON-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: ps459
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..8834
; FEATURE:
; NAME/KEY: intron
; LOCATION: 8868..10014
; FEATURE:
; NAME/KEY: intron
```

;> LOCATION: 10511..12277  
;> FEATURE: exon  
;> NAME/KEY: exon  
;> LOCATION: 8835..8867  
;> FEATURE:  
;> NAME/KEY: exon  
;> LOCATION: 10015..10510  
;> FEATURE:  
;> NAME/KEY: exon  
;> LOCATION: 12278..12443  
US-08-462-437-4

Query Match 78.1%; Score 16.4; DB 4; Length 13104;  
Best Local Similarity 94.4%; Pred. No. 14;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 ggggtggaggctatttgg 21  
||||| |||||||  
Db 10242 GGGTGGTGCTATTGG 10225

RESULT 11  
US-09-111-470-7  
;> Sequence 7, Application US/09111470  
;> Patent No. 6277959  
;> GENERAL INFORMATION:  
;> APPLICANT: Valladeau, Jenny  
;> APPLICANT: Ravel, Odile  
;> APPLICANT: Bates, Elizabeth E.M.  
;> APPLICANT: Ford, John  
;> APPLICANT: Saeland, Sem  
;> APPLICANT: Lebecque, Serge J.E.  
;> TITLE OF INVENTION: Mammalian Membrane Protein Genes;  
;> TITLE OF INVENTION: Related Reagents  
;> NUMBER OF SEQUENCES: 11  
;> CORRESPONDENCE ADDRESS:  
;> ADDRESSEE: DNAX Research Institute  
;> STREET: 901 California Avenue  
;> CITY: Palo Alto  
;> STATE: California  
;> COUNTRY: USA  
;> ZIP: 94304-1104  
;> COMPUTER READABLE FORM:  
;> MEDIUM TYPE: Floppy disk  
;> COMPUTER: IBM PC compatible  
;> OPERATING SYSTEM: PC-DOS/MS-DOS  
;> SOFTWARE: Patentin Release #1.0, Version #1.30  
;> CURRENT APPLICATION DATA:  
;> APPLICATION NUMBER: US/09/111.470  
;> FILING DATE: 08-JUL-1998  
;> CLASSIFICATION:  
;> PRIOR APPLICATION DATA:  
;> APPLICATION NUMBER: US 60/053,080  
;> FILING DATE: 09-JUL-1997  
;> ATTORNEY/AGENT INFORMATION:  
;> NAME: Ching, Edwin P.  
;> REGISTRATION NUMBER: 34,090  
;> REFERENCE/DOCKET NUMBER: SF0695  
;> TELECOMMUNICATION INFORMATION:  
;> TELEPHONE: (650)852-9196  
;> TELEFAX: (650)496-1200  
;> INFORMATION FOR SEQ ID NO: 7:  
;> SEQUENCE CHARACTERISTICS:  
;> LENGTH: 1418 base pairs  
;> TYPE: nucleic acid  
;> STRANDEDNESS: single  
;> TOPOLOGY: linear  
;> MOLECULE TYPE: cDNA  
;> FEATURE:  
;> NAME/KEY: CDS  
;> LOCATION: 279..992  
;> FEATURE:

;> NAME/KEY: misc\_feature  
;> LOCATION: 1348  
;> OTHER INFORMATION: /note= "poly-A addition motif"  
US-09-111-470-7

Query Match 77.1%; Score 16.2; DB 4; Length 1418;  
Best Local Similarity 85.7%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atagggtggaggctatttgg 21  
||||| |||||||  
Db 602 AAAGGATTGGAGGCTATTGG 622

RESULT 12  
US-08-147-777-3  
;> Sequence 3, Application US/08147777  
;> Patent No. 5914265  
;> GENERAL INFORMATION:  
;> APPLICANT: Roop, Dennis R.  
;> APPLICANT: Rothnagel, Joseph A.  
;> APPLICANT: Greenhalgh, David A.  
;> APPLICANT: Yuspa, Stuart H.  
;> TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS  
;> TITLE OF INVENTION: AND METHODS OF USE  
;> NUMBER OF SEQUENCES: 5  
;> CORRESPONDENCE ADDRESS:  
;> ADDRESSEE: LYON & LYON  
;> STREET: 611 West Sixth Street  
;> CITY: Los Angeles  
;> STATE: California  
;> COUNTRY: U.S.A.  
;> ZIP: 90017  
;> COMPUTER READABLE FORM:  
;> MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
;> COMPUTER: IBM PC compatible  
;> OPERATING SYSTEM: IBM MS-DOS (Version 5.0)  
;> SOFTWARE: WordPerfect (Version 5.1)  
;> CURRENT APPLICATION DATA:  
;> APPLICATION NUMBER: US/08/147,777  
;> FILING DATE:  
;> CLASSIFICATION: 800  
;> PRIOR APPLICATION DATA:  
;> PRIOR APPLICATION DATA: including application  
;> PRIOR APPLICATION DATA: described below: two  
;> APPLICATION NUMBER: 07/876,289  
;> FILING DATE: April 30, 1992  
;> APPLICATION NUMBER: Unassigned (204/144)  
;> FILING DATE: October 29, 1993  
;> ATTORNEY/AGENT INFORMATION:  
;> NAME: Warburg, Richard J.  
;> REGISTRATION NUMBER: 32,327  
;> REFERENCE/DOCKET NUMBER: 204/153  
;> TELECOMMUNICATION INFORMATION:  
;> TELEPHONE: (213) 489-1600  
;> TELEFAX: (213) 955-0440  
;> TELEX: 67-3510  
;> INFORMATION FOR SEQ ID NO: 3:  
;> SEQUENCE CHARACTERISTICS:  
;> LENGTH: 24979 base pairs  
;> TYPE: nucleic acid  
;> STRANDEDNESS: single  
;> TOPOLOGY: linear  
;> MOLECULE TYPE: DNA (genomic)  
US-08-147-777-3

Query Match 77.1%; Score 16.2; DB 2; Length 24979;  
Best Local Similarity 85.7%; Pred. No. 19;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atagggtggaggctatttgg 21

```
Db 3248 AAAGGTGGGAAGCTATATGG 3268
|||||
RESULT 13
US-08-452-872-3
; Sequence 3, Application US/08452872
; Patent No. 6057298
; GENERAL INFORMATION:
; APPLICANT: Roop, Dennis R.
; APPLICANT: Rothnagel, Joseph A.
; APPLICANT: Greenhalgh, David A.
; APPLICANT: Yuspa, Stuart H.
; TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,872
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,777
; FILING DATE:
; APPLICATION NUMBER: 07/876,289
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: Unassigned (204/144)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/153
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24979 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-452-872-3

Query Match 77.1%; Score 16.2; DB 3; Length 24979;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
| |||||
Db 3248 AAAGGTGGGAAGCTATATGG 3268

RESULT 14
US-08-136-277-23
; Sequence 23, Application US/08136277
; Patent No. 5644045
; GENERAL INFORMATION:
; APPLICANT: MANDEL, Jean-Louis
; APPLICANT: AUBOURG, Patrick
; APPLICANT: MOSSER, Jean
; APPLICANT: SARDE, Claude
; TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
; TITLE OF INVENTION: CORRESPONDING PROTEIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

PCT-US93-03985-3

Query Match 77.1%; Score 16.2; DB 5; Length 24979;
Best Local Similarity 85.7%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atagggtggaggctatttgg 21
| |||||
Db 3248 AAAGGTGGGAAGCTATATGG 3268

RESULT 15
US-08-136-277-23
; Sequence 23, Application US/08136277
; Patent No. 5644045
; GENERAL INFORMATION:
; APPLICANT: MANDEL, Jean-Louis
; APPLICANT: AUBOURG, Patrick
; APPLICANT: MOSSER, Jean
; APPLICANT: SARDE, Claude
; TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
; TITLE OF INVENTION: CORRESPONDING PROTEIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

PCT-US93-03985-3
```



```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136,277
; FILING DATE: 15-OCT-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: B2272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1025 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-136-277-23

```

```

Query Match      75.2%; Score 15.8; DB 1; Length 1025;
Best Local Similarity 89.5%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 2 taggggtggaggctatttg 20
   ||| ||||| ||| ||
Db 148 TAGGTTGGGAGGCTATGTG 166

```

Search completed: February 25, 2002, 18:05:11  
Job time: 18559 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:20:51 ; Search time 8261.74 Seconds  
(without alignments)  
27.314 Million cell updates/sec

Title: US-09-698-903B-3

Perfect score: 21

Sequence: 1 atagggtggaggctatttgg 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estfun:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estom:\*

5: em\_estpl:\*

6: em\_estba:\*

7: em\_estro:\*

8: em\_estov:\*

9: em\_htc:\*

10: gb\_est1:\*

11: gb\_est2:\*

12: gb\_htc:\*

13: gb\_gss:\*

14: em\_gss\_fun:\*

15: em\_gss\_hum:\*

16: em\_gss\_inv:\*

17: em\_gss\_pln:\*

18: em\_gss\_pro:\*

19: em\_gss\_rod:\*

20: em\_gss\_vrt:\*

21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	Description
c 1	17.4	82.9	212 11 BG983367
c 2	17.4	82.9	429 11 BF924973
c 3	17.4	81.0	501 10 AW638905
c 4	17.4	81.0	583 10 AW635936
c 5	16.8	80.0	285 10 AV145846
c 6	16.8	80.0	321 11 F05693
c 7	16.8	80.0	424 11 W57725
c 8	16.8	80.0	429 11 BI188834
c 9	16.8	80.0	587 11 BG494856
c 10	16.8	80.0	626 10 AW644184
c 11	16.8	80.0	667 10 AW639842
c 12	16.8	80.0	892 11 BF167952

c 13	16.8	80.0	914	13	CNS03N93
c 14	16.4	78.1	284	13	AZ016369
c 15	16.4	78.1	284	11	BG210286
c 16	16.4	78.1	385	10	AW736851
c 17	16.4	78.1	462	11	R41615
c 18	16.4	78.1	575	10	AW578936
c 19	16.4	78.1	658	13	BH036913
c 20	16.4	78.1	714	11	BG217189
c 21	16.4	78.1	728	11	BG204492
c 22	16.4	78.1	781	13	AZ188345
c 23	16.4	78.1	798	11	BG199886
c 24	16.4	78.1	822	11	BG194262
c 25	16.4	78.1	829	11	BG199887
c 26	16.4	78.1	837	11	BG211849
c 27	16.4	78.1	838	11	BG209779
c 28	16.4	78.1	850	11	BG215054
c 29	16.4	78.1	859	11	BG219740
c 30	16.4	78.1	866	11	BG205014
c 31	16.4	78.1	962	11	BG195797
c 32	16.2	77.1	162	10	BE011473
c 33	16.2	77.1	166	11	BF328785
c 34	16.2	77.1	228	10	AV338030
c 35	16.2	77.1	263	10	AA178060
c 36	16.2	77.1	265	10	BB252853
c 37	16.2	77.1	270	10	BB307020
c 38	16.2	77.1	274	10	BB008234
c 39	16.2	77.1	298	10	BB464468
c 40	16.2	77.1	321	10	AA850434
c 41	16.2	77.1	323	10	BB319109
c 42	16.2	77.1	326	10	BB319885
c 43	16.2	77.1	326	10	BB461095
c 44	16.2	77.1	348	13	AQ610088
c 45	16.2	77.1	385	10	BE113833

ALIGNMENTS

RESULT	1
LOCUS	BG983367/c
DEFINITION	IL5-CN0067-060301-384-fl2 CN0067 Homo sapiens cdna, mRNA sequence.
ACCESSION	BG983367
VERSION	BG983367.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brungshein, A., de Oliveira, P., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil. Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-CN0067-060301-fl2&t3=2001-03-06&t4=1) Seq primer: puc 18 forward

High quality sequence stop: 211.

FEATURES  
source  
1. .212  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NC0067"  
/dev\_stage="Adult"  
/note="Organ: colon\_normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
38 a 84 c 32 g 58 t

BASE COUNT  
ORIGIN  
Query Match 82.9%; Score 17.4; DB 11; Length 212;  
Best Local Similarity 94.7%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 2 taggggtggaggctatttg 20  
|||||  
Db 108 TAGGGTGGGAGGCTATGTG 90

# RESULT 2

LOCUS BF924973 429 bp mRNA EST 19-JAN-2001  
DEFINITION IL2-NT0200-281100-263-C11 NT0200 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF924973  
VERSION BF924973.1 GI:12320948  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 429)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&t=IL2-NT0200-  
281100-263-C11&t3=2000-11-28&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 351.  
FEATURES  
source  
1. .429  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NT0200"  
/dev\_stage="Adult"  
/note="Organ: nervous\_tumor; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
99 a 106 c 132 g 90 t 2 others

Query Match 82.9%; Score 17.4; DB 11; Length 429;  
Best Local Similarity 94.7%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Caps 0;  
Qy 3 aggggtggaggctatttg 21  
|||||  
Db 392 AGCGTGGGAGGCTATTTGG 410

# RESULT 3

LOCUS AW638905/c 501 bp mRNA EST 26-APR-2001  
DEFINITION bl76B06.w1 Blackshear/Soares normalized Xenopus egg library Xenopus  
laevis cDNA clone PBX0076B06 5', mRNA sequence.  
ACCESSION AW638905  
VERSION AW638905.1 GI:7396075  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 501)  
Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.,  
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman  
J.W., Bonaldo,M.F. and Soares,M.B.  
The NIH's Xenopus maternal EST project: interim analysis of the  
first 13,879 ESTs from unfertilized eggs  
Gene 267 (1), 71-87 (2001)  
21211403

CONTACT: Perry J. Blackshear  
Office of Clinical Research and Laboratory of Signal Transduction  
National Institute of Environmental Health Sciences  
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,  
USA  
Tel: 919 541-4899  
Fax: 919 541-4571  
Email: black009@niehs.nih.gov  
Clone is available through Research Genetics, Inc., 2130 Memorial  
Parkway, Huntsville, AL 35901  
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email  
cdna@resgen.com  
DNA Sequencing and analyses performed by National Institutes of  
Health Intramural Sequencing Center (NISC).  
PCR PRIMERS  
FORWARD: TGTAACGACGCGCCAGT  
BACKWARD: CAGGAAACAGCTATGACC  
Plate: 0076 row: B column: 06  
Seq primer: T7 primer.

FEATURES  
source  
1. .501  
Location/Qualifiers  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="PBX0076B06"  
/clone\_lib="Blackshear/Soares normalized Xenopus egg  
library"  
/sex="female"  
/tissue\_type="unfertilized egg"  
/cell\_type="unfertilized egg"  
/dev\_stage="unfertilized egg"  
/lab\_host="DH10B"  
/note="Vector: pT7T3-Pac; Site\_1: EcoRI; Site\_2: NotI;  
PolyA-selected mRNA was prepared from unfertilized Xenopus  
laevis eggs. The library was constructed in the vector  
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and  
Soares, M.B. Normalization and subtraction: two

approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dri18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 X 10<sup>5</sup> recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT  
ORIGIN

161 a 116 c 106 g 118 t

Query Match 81.0%; Score 17; DB 10; Length 501;  
Best Local Similarity 100.0%; Pred. No. 9.3e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 taggggtggaggctatt 18  
|||||

Db 317 TAGGGTGGGAGGCTATT 301

RESULT 4

AW635936/c

LOCUS

AW635936 583 bp mRNA EST 26-APR-2001

DEFINITION

bl140d03.w1 Blackshear/Soares normalized Xenopus egg library Xenopus

laevis cDNA clone PBX0040D03 5', mRNA sequence.

ACCESSION

AW635936

VERSION

AW635936.1 GI:7393017

KEYWORDS

EST,

SOURCE

African clawed frog.

ORGANISM

Xenopus laevis

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 583)

Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.

Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman

, J.W., Bonaldo, M.F. and Soares, M.B.

The NIEHS Xenopus maternal EST project: Interim analysis of the

first 13,879 ESTs from unfertilized eggs

Gene 267 (1), 71-87 (2001)

21211403

Contact: Perry J. Blackshear

Office of Clinical Research and Laboratory of Signal Transduction

National Institute of Environmental Health Sciences

A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,

USA

Tel: 919 541-4899

Fax: 919 541-4571

Email: black009@niehs.nih.gov

Clone is available through Research Genetics, Inc., 2130 Memorial

Parkway, Huntsville, AL 35901

phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email

cdna@resgen.com

DNA sequencing and analyses performed by National Institutes of

Health Intramural Sequencing Center (NISC).

PCR Primers

FORWARD: TGTAACACGACGGCCAGT

BACKWARD: CAGGAACAGCTATGACC

Plate: 0040 row: D column: 03

Seq primer: T7 primer.

Location/Qualifiers

1..583

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="PBX0040D03"

/clone\_lib="Blackshear/Soares normalized Xenopus egg

library"

/sex="female"

/tissue\_type="unfertilized egg"

/cell\_type="unfertilized egg"

/dev\_stage="unfertilized egg"

/lab\_host="DH10B"

/note="Vector: pT7T3-Pac; Site\_1: EcoRI; Site\_2: NotI;

FEATURES  
source

PolyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dri18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 X 10<sup>5</sup> recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT  
ORIGIN

185 a 126 c 128 g 144 t

Query Match 81.0%; Score 17; DB 10; Length 583;  
Best Local Similarity 100.0%; Pred. No. 9.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 taggggtggaggctatt 18  
|||||

Db 433 TAGGGTGGGAGGCTATT 417

RESULT 5

AV145846/c

LOCUS

AV145846

DEFINITION

clone 2810449H10, mRNA sequence.

ACCESSION

AV145846

VERSION

AV145846.1 GI:5349979

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 285)

Carinci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,

Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara

, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,

Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,

Sato, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Sugahara

, Y., Suzuki, H., Suzuki, H., Taten, M., Tomaru, Y., Tomimaga, N.,

Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,

Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp

Thermolabile and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

feature details.

Location/Qualifiers

1..285

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="2810449H10"

/clone\_lib="Mus musculus C57BL/6J 10-11 day embryo"

/sex="mixed"

/dev\_stage="10-11 day embryo"

BASE COUNT 63 a 87 c 53 g 82 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 285;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggggtggaggctattgg 21  
 |||||  
 Db 47 TAGGGTGGGAGCAATTAGG 28

RESULT 6  
 LOCUS F05693/c  
 DEFINITION HSC0JC091 normalized infant brain cDNA Homo sapiens cDNA clone  
 C-0jc09, mRNA sequence.  
 ACCESSION F05693  
 VERSION F05693.1 GI:669509  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 321)  
 AUTHORS Aufiray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes  
 , M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F.,  
 Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y.,  
 Sebastiani-Kabaktchis, C. and Tessier, A.  
 TITLE IMAGE: molecular integration of the analysis of the human genome  
 and its expression  
 JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)  
 MEDLINE 95277334  
 COMMENT Contact: Genethon  
 Genexpress-Genethon  
 Genethon Centre de recherche sur le Genome Humain  
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE  
 Tel: 33169472800  
 Fax: 33160778698  
 Email: genexpress@genethon.fr  
 Single read.  
 Genexpress\_library\_id: C; Genexpress\_sequence\_id: ylc-0jc09  
 Seq primer: (-21)M13-universal.

FEATURES  
 source  
 1..321  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="c-0jc09"  
 /clone\_lib="normalized infant brain cDNA"  
 /sex="Female"  
 /tissue\_type="total brain"  
 /dev\_stage="3 months old"  
 /note="Organ: brain; Vector: lafmid BA; Site:1: HindIII;  
 Site\_2: NotI; sex=Female; dev\_stage=3 months old;  
 isolate=muscular atrophy patient; tissue\_type=total brain  
 ; total mRNA was oligo-(GT) primed and directionally  
 cloned 5' -> 3' into the HindIII -> NotI sites of the  
 lafmid BA vector. Clone library from B.Souares, Psychiatry  
 Dept. Columbia University, USA. Normalization\_method:  
 Bento Soares, P.N.A.S in press"  
 BASE COUNT 108 a 53 c 46 g 114 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 321;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggggtggaggctattgg 21  
 |||||  
 Db 106 TGGGGTGGGAGCAATTGG 87

RESULT 7  
 W57725/c

LOCUS W57725 424 bp mRNA EST 15-OCT-1996  
 DEFINITION zd20c10.sl Soares\_fetal\_heart\_NbHH19W Homo sapiens cDNA clone  
 IMAGE:341202 3', mRNA sequence.  
 ACCESSION W57725  
 VERSION W57725.1 GI:1364441  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 424)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
 , M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston  
 , R., Williamson, A., Wohldmann, P. and Wilson, R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Wilton RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 866 Std Error: 0.00  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 376.  
 Location/Qualifiers  
 1..424  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1266577"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:341202"  
 /clone\_lib="Soares\_fetal\_heart\_NbHH19W"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a  
 modified polylinker; Site:1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGGAGCGGCATCTTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT73 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by  
 M.Fatima Bonaldo. This library was constructed from the  
 same fetus as the fetal lung library, Soares fetal lung  
 NbHH19W."

BASE COUNT 145 a 87 c 77 g 115 t  
 ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 424;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atagggtggaggctattgg 20  
 |||||  
 Db 374 ATGGGTGGGAGGCTCTTG 355

RESULT 8  
 BI188834/c

LOCUS BI188834 429 bp mRNA EST 10-JUL-2001  
 DEFINITION d3d12fs.f1 Fusarium sporotrichioides Tri 10 overexpressed cDNA  
 library Fusarium sporotrichioides cDNA clone d3d12fs 3', mRNA  
 sequence.  
 ACCESSION BI188834  
 VERSION BI188834.1 GI:14662513  
 KEYWORDS EST.  
 SOURCE Fusarium sporotrichioides.

```

ORGANISM  Fusarium sporotrichioides
REFERENCE  1 (bases 1 to 429)
AUTHORS    Ren, Q., Tag, A., Peplow, A., Lai, H., Kupfer, C., Peterson, A., Beremand
           , M. and Roe, B.
TITLE      Analysis of a Fusarium sporotrichioides EST database
JOURNAL    Unpublished (2001)
COMMENT    Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
           Department of Chemistry and Biochemistry
           Advanced Center for Genome Technology, University of Oklahoma
           620 Parrington Oval, Norman, OK 73019, USA
           Tel: 405 325 4912
           Fax: 405 325 7762
           Email: broe@ou.edu
           Contact Dr. Marian Beremand regarding clone availability. Included
           is the best homolog from a blastx search of Genbank nr 04-09-01
           71.3.9.gi|3550690|emb|CAA09 (AJ010199) variable surface
           glycoprotein.
Seq primer: M13-20
High quality sequence stop: 405.
Location/Qualifiers
1..429
/organism="Fusarium sporotrichioides"
/strain="Tri 10"
/db_xref="taxon:5514"
/clone="d3d12fs"
/clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
cDNA library"
/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT  137 a 102 c 68 g 122 t
ORIGIN
1 atagggtggaggctatttg 20
|||||
Db 164 ATTGGAGGAGGCTATTG 145

Query Match      80.0%; Score 16.8; DB 11; Length 429;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT  9
LOCUS    BG494856      587 bp      mRNA      EST      27-MAR-2001
DEFINITION  602540862p1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4671903 5',
mRNA sequence.
ACCESSION  BG494856
VERSION    BG494856.1 GI:13456371
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 587)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1487 row: 1 column: 16
            High quality sequence stop: 584.
Location/Qualifiers

```

```

1..587
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4671903"
/clone_lib="NIH_MGC_59"
/tissue_type="mucoepidermoid carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctcgcc); Site_2: SfiI (ggccattggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTATGGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGAGCGGCACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT  197 a 104 c 87 g 199 t
ORIGIN
1 taggggtggaggctatttg 21
|||||
Db 191 TGGGTGGGAGCAATTGG 172

Query Match      80.0%; Score 16.8; DB 11; Length 587;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT  10
LOCUS    AW644184      626 bp      mRNA      EST      26-APR-2001
DEFINITION  cm38d02.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0137D02 5', mRNA sequence.
ACCESSION  AW644184
VERSION    AW644184.1 GI:7401561
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 626)
AUTHORS    Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.,
            Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman
            , J.W., Bonaldo, M.F. and Soares, M.B.
TITLE      The NIEHS Xenopus maternal EST project: interim analysis of the
            first 13,879 ESTs from unfertilized eggs
JOURNAL    Gene 267 (1), 71-87 (2001)
MEDLINE    21211403
COMMENT    Contact: Perry J. Blackshear
            Office of Clinical Research and Laboratory of Signal Transduction
            National Institute of Environmental Health Sciences
            A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
            USA
            Tel: 919 541-4899
            Fax: 919 541-4571
            Email: black009@niehs.nih.gov
            Clone is available through Research Genetics, Inc., 2130 Memorial
            Parkway, Huntsville, AL 35901
            phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
            cdna@resgen.com
            DNA Sequencing and analyses performed by National Institutes of
            Health Intramural Sequencing Center (NISC).
            PCR Primers
            FORWARD: TGTAACGACGCGCCAGT
            BACKWARD: CAGGAACAGCTATGACC
            Plate: 0137 row: D column: 02
            Seq primer: T7 primer.
Location/Qualifiers

```

## source

```

1. .626
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0137D02"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      204 a 116 c 131 g 175 t
ORIGIN

Query Match      80.0%; Score 16.8; DB 10; Length 626;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggtggaggctattgg 21
|||||
Db 131 TAGGGTGGGAGGATATTAGG 150

RESULT 11
AW639842
LOCUS      AW639842      667 bp      mRNA      EST      26-APR-2001
DEFINITION b187hl1.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
laevis cDNA clone PBX0087H11 5', mRNA sequence.
ACCESSION  AW639842
VERSION     AW639842.1 GI:7397045
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 667)
Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
J.W., Bonaldo,M.F. and Soares,M.B.
the NIEHS Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR PRIMERS
FORWARD: TGTAACACGCGCCAGT

```

```

BACKWARD: CAGGAACACGCTATGACC
Plate: 0087 row: H column: 11
Seq primer: T7 primer
FEATURES
source
Location/Qualifiers
1. .667
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="PBX0087H11"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/notes="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT      214 a 124 c 144 g 185 t
ORIGIN

```

```

Query Match      80.0%; Score 16.8; DB 10; Length 667;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 taggtggaggctattgg 21
|||||
Db 143 TAGGGTGGGAGGATATTAGG 162

RESULT 12
BF167952/c
LOCUS      BF167952      892 bp      mRNA      EST      30-OCT-2000
DEFINITION 601775688F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:4017644 5',
mRNA sequence.
ACCESSION  BF167952
VERSION     BF167952.1 GI:11048304
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 892)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9266 row: o column: 21
High quality sequence stop: 662.
Location/Qualifiers
1. .892
/organism="Mus musculus"
/strain="C57BL/6J (f1)"
/db_xref="taxon:10090"
/clone="IMAGE:4017644"
FEATURES
source
Location/Qualifiers

```



```

/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site.1: Sali;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT      241 a   225 c   237 g   189 t
ORIGIN

Query Match      80.08; Score 16.8; DB 11; Length 892;
Best Local Similarity 90.08; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 taggggtggaggctattgg 21
||||| ||||||| |||
Db  871 TAGGGTGTGAGGCATATGG 852

RESULT 13
CNS03N93/c
LOCUS      914 bp      DNA      GSS      17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
039K16 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  AL251760.1 GI:797272
VERSION     AL251760.1
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM    Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 914)
AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Fitzames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 914)
AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fitzames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE      Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL    Unpublished
REFERENCE   3 (bases 1 to 914)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT    This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES             Location/Qualifiers
     source           1..914
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="039K16"
                     /clone_lib="G"
BASE COUNT      202 a   257 c   239 g   212 t   4 others
ORIGIN

Query Match      80.08; Score 16.8; DB 13; Length 914;
Best Local Similarity 85.78; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site.1: Sali;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT      241 a   225 c   237 g   189 t
ORIGIN

Query Match      80.08; Score 16.8; DB 11; Length 892;
Best Local Similarity 90.08; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 taggggtggaggctattgg 21
||||| ||||||| |||
Db  871 TAGGGTGTGAGGCATATGG 852

RESULT 14
AZ016369
LOCUS      184 bp      DNA      GSS      25-FEB-2000
DEFINITION RPCI-23-264M11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-264M11
, DNA sequence.
ACCESSION  AZ016369
VERSION     AZ016369.1 GI:7091753
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 184)
AUTHORS    Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE      Mouse BAC End Sequences from Library RPCI-23
JOURNAL    Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-264M11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 264 row: M column: 11
Seg primer: Sp6
Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..184
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-264M11"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site.1:
EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT      40 a   36 c   49 g   59 t
ORIGIN

Query Match      78.18; Score 16.4; DB 13; Length 184;
Best Local Similarity 94.48; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  3 aggggtggaggctattgg 20
||||| ||||||| |||
Db  60 AGGGTGGGAGGCAATTGG 77

RESULT 15
BG210286/c
LOCUS      284 bp      mRNA      EST      21-APR-2001
DEFINITION RST29821 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.

```

ACCESSION BG210286  
VERSION BG210286.1 GI:13731973  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,  
Lerner,L., Krashinsky,D., McElligott,K., Clark,S., Mays,R., Smith,E.,  
Veloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,  
and Ducar,M.  
TITLE Creation of Genome-wide Protein Expression Libraries using Random  
JOURNAL Activation of Gene Expression  
COMMENT Nat. Biotechnol. 19 (5), 440 (2001) in press  
Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scaine@atersys.com  
High quality sequence stop: 234.  
FEATURES  
Source  
1..284  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Athersys RAGE Library"  
/cell\_line="HT1080"  
/note="See 'Creation of Genome-wide Protein Expression  
Libraries using Random Activation of Gene Expression',  
Nature Biotechnology, in press. Note that even though the  
cell type indicated is HT1080, since a random activation  
method was used, these sequence tags are not necessarily  
expressed in HT1080 under normal circumstances."  
BASE COUNT 88 a 82 c 43 g 71 t  
ORIGIN

Query Match 78.1% Score 16.4; DB 11; Length 284;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 4 ggggtggaggctatttgg 21  
||||| |||||||||  
Db 238 GGGTGGGTGGCTATTGG 221

Search completed: February 25, 2002, 17:20:54  
Job time: 16147 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:06 ; Search time 2331.3 Seconds  
(without alignments)  
106.146 Million cell updates/sec

Title: US-09-698-903B-4

Perfect score: 15

Sequence: 1 ntcgastwtsgwgtt 15

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Match	Length	DB	ID	Description
1	12.4	82.7	15	6	AX046662	AX046662 Sequence
2	12.4	82.7	15	6	AX063588	AX063588 Sequence
3	12.4	82.7	15	6	AX127751	AX127751 Sequence
4	12.4	82.7	15	6	AX167666	AX167666 Sequence
5	12.4	82.7	15	6	E08065	E08065 Non-specific
6	12.4	82.7	206	10	MMU0260	U08260 Mus musculu
7	12.4	82.7	240	8	GMU26701	U26701 Glycine max
8	12.4	82.7	411	8	PUMPAL2A	L37355 Petroselinu
9	12.4	82.7	415	6	AX127755	AX127755 Sequence
10	12.4	82.7	636	9	HSWCRP1	AF073484 Homo sapi
11	12.4	82.7	676	3	AF283269	AF283269 Anopheles
12	12.4	82.7	733	3	GRO251757	AJ251757 Globodera
13	12.4	82.7	905	8	GCU94186	U94186 Colletotric
14	12.4	82.7	927	8	AF049922	AF049922 Petunia x
15	12.4	82.7	942	8	AF328994	AF328994 Arabidops
16	12.4	82.7	1088	8	AF162910	AF162910 Hirschfel
17	12.4	82.7	1092	8	AF161333	AF161333 Raphanus
18	12.4	82.7	1092	8	AF162913	AF162913 Raphanus
19	12.4	82.7	1321	6	AX046673	AX046673 Sequence
20	12.4	82.7	1499	3	U24436	U24436 Mayetiola d
21	12.4	82.7	2115	8	CYR242496	AJ242496 Candida t
22	12.4	82.7	2154	8	AF052582	AF052582 Fusarium
23	12.4	82.7	2246	8	AF387019	AF387019 Arabidops
24	12.4	82.7	2275	3	AF218866	AF218866 Caenorhab
25	12.4	82.7	2384	8	PCPAL2	X81158 P.crispum m
26	12.4	82.7	2427	9	AF191338	AF191338 Homo sapi
27	12.4	82.7	2851	10	RATAP0A4G	M13508 Rat apolipo
28	12.4	82.7	2914	9	AK022864	AK022864 Homo sapi
29	12.4	82.7	2931	8	SCYGR065C	Z72850 S.cerevisia
30	12.4	82.7	3084	3	AF395331	AF395331 Drosophil
31	12.4	82.7	3246	3	AF245516	AF245516 Drosophil
32	12.4	82.7	3282	2	AC019765	AC019765 Drosophil
33	12.4	82.7	3614	3	DME131630	AJ131630 Drosophil
34	12.4	82.7	3855	10	BC005506	BC005506 Mus muscu
35	12.4	82.7	3887	3	AF185069	AF185069 Leishmani
36	12.4	82.7	4233	6	AX017147	AX017147 Sequence
37	12.4	82.7	4271	9	AB029020	AB029020 Homo sapi
38	12.4	82.7	4501	14	MYU88057	U88057 Meiao virus
39	12.4	82.7	4953	4	RABGLCOGEN	L10605 Oryctolagus
40	12.4	82.7	5288	2	AC019630	AC019630 Drosophil
41	12.4	82.7	5315	8	WHTGIR	J02817 Wheat gibbe
42	12.4	82.7	6201	2	AC018291	AC018291 Drosophil
43	12.4	82.7	6211	3	AF325859	AF325859 Choriston
44	12.4	82.7	7101	3	AF361883	AF361883 Schlstoso
45	12.4	82.7	9276	2	AC014293	AC014293 Drosophil

ALIGNMENTS

RESULT 1

AX046662

LOCUS AX046662 15 bp DNA

DEFINITION Sequence 8 from Patent WO0068403.

ACCESSION AX046662

VERSION AX046662.1 GI:11876228

KEYWORDS .

SOURCE synthetic construct.

ORGANISM synthetic construct.

REFERENCE 1 (bases 1 to 15)

AUTHORS Paul, W.B., Scott, R.J., Hird, D.U. and Hodge, R.U.

TITLE Tapetum-specific promoters

JOURNAL Patent: WO 0068403-A 8 16-NOV-2000;

KEYWORDS BIOGENMA UK LIMITED (GB)

FEATURES Location/Qualifiers

source . .15

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="Primer"

```

misc_feature 1
/note="n is a, t, g, or c"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 2
AX063588 15 bp DNA PAT 24-JAN-2001
LOCUS
DEFINITION Sequence 7 from Patent WO0100801.
ACCESSION AX063588
VERSION AX063588.1 GI:12541312
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 15)
AUTHORS Habu,Y., mittelsten Scheid,O., Amedeo,P. and Paszkowski,J.
TITLE Gene involved in epigenetic gene silencing
JOURNAL Patent: WO 0100801-A 7 04-JAN-2001;
Novartis AG (CH) ; Novartis-Erfindungen Verwaltungsgesellschaft
m.b.H. (AT)
FEATURES
source
Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Synthetic Oligonucleotide"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 3
AX127751 15 bp DNA PAT 15-MAY-2001
LOCUS
DEFINITION Sequence 4 from Patent WO0131042.
ACCESSION AX127751
VERSION AX127751.1 GI:14134398
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 15)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 4 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer MDB285"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 4
AX167666 15 bp DNA PAT 03-JUL-2001
LOCUS
DEFINITION Sequence 11 from Patent WO0144277.
ACCESSION AX167666
VERSION AX167666.1 GI:14597053
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 15)
AUTHORS Wegrich Glover,L., Budziszewski,G.J., Levin,J.Z. and Zhou,Q.
TITLE Herbicide target genes and methods
JOURNAL Patent: WO 0144277-A 11 21-JUN-2001;
Syngenta Participations AG (CH)
FEATURES
source
Location/Qualifiers
1..15
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide"
BASE COUNT 1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
|||||
Db 2 TCGASTWTSGWGT 15

RESULT 5
E08065 15 bp DNA PAT 29-SEP-1997
LOCUS
DEFINITION Non-specific primer.
ACCESSION E08065
VERSION E08065.1 GI:2176190
KEYWORDS JP 1994253843-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 15)
AUTHORS Riyuu,Y., Robaato,E.U. and Mitsukawa,N.
TITLE METHOD FOR AMPLIFYING DNA SEQUENCE
JOURNAL Patent: JP 1994253843-A 4 13-SEP-1994;
CHIKU KANKYO SANGYO GIJUTSU KENKYU KIKO, MITSUI GIYOUSAI
SHOKUBUTSU BIO KENKYUSHO:KK
COMMENT
OS None
OC Artificial sequences.
PN JP 1994253843-A/4
PD 13-SEP-1994
PF 03-MAR-1993 JP 1993043006
PI RIYUU YOKOU, ROBAATO EFU UITSUTEIA, MITSUKAWA NORIHIRO PC
C12N15/10,C12Q1/68;
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FH source
1..15
/organism="Artificial sequences",

```

```
FEATURES
  source
    Location/Qualifiers
      1. .15
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT      1 a 1 c 3 g 5 t 5 others
ORIGIN

Query Match      82.7%; Score 12.4; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
    |||||:|||||
Db 2 TCGASTWTSGWGT 15

RESULT 6
MMU80260
LOCUS      MMU80260      206 bp      DNA      ROD      19-DEC-1996
DEFINITION Mus musculus clone ma9 L1 retrotransposon LINE1 repeat region.
ACCESSION U80260
VERSION   U80260.1 GI:11737239
KEYWORDS  house mouse.
SOURCE    house mouse.
ORGANISM  Mus musculus
REFERENCE 1 (bases 1 to 206)
AUTHORS  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL   Woodcock,D.M. and Linzenmeyer,M.E.
          Direct Submission
          Submitted (29-NOV-1996) Research, Peter MacCallum Cancer Institute,
          St. Andrew's Place, East Melbourne, Vic 3000, Australia
FEATURES  Location/Qualifiers
  source
    1. .206
      /organism="Mus musculus"
      /transposon="L1 retrotransposon"
      /db_xref="taxon:10090"
      /clone="ma9"
repeat_region <1..>206
BASE COUNT      32 a 14 c 74 g 86 t
ORIGIN

Query Match      82.7%; Score 12.4; DB 10; Length 206;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
    |||||:|||||
Db 45 TCGAGTTTCGAGTT 58

RESULT 7
GMU26701/c
LOCUS      GMU26701      240 bp      DNA      PLN      26-JUL-2001
DEFINITION Glycine max satellite STR120-B.1.
ACCESSION U26701
VERSION   U26701.1 GI:1142703
KEYWORDS  soybean.
SOURCE    Glycine max
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
          Glycine.
          1 (bases 1 to 240)
REFERENCE Morgante,M., Jurman,I., Shi,L., Zhu,T., Keim,P. and Rafalski,J.A.
AUTHORS   The STR120 satellite DNA of soybean: organization, evolution and
TITLE     Chromosomal specificity
JOURNAL   Chromosome Res. 5 (6), 363-373 (1997)
MEDLINE   98030220

FEATURES  Location/Qualifiers
  source
    1. .411
      /organism="Petroselinum crispum"
      /db_xref="taxon:4043"
      /note="(vector lambda EMBL4)"
promoter      1. .260
TATA_signal   230. .235
gene          409. .411
            /gene="PAL-2"
BASE COUNT      128 a 99 c 40 g 144 t
ORIGIN

Query Match      82.7%; Score 12.4; DB 8; Length 411;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
    |||||:|||||
Db 363 TCGAGTTTCGAGTT 376

RESULT 9
AX127755
LOCUS      AX127755      415 bp      DNA      PAT      15-MAY-2001

FEATURES  Location/Qualifiers
  source
    1. .240
      /organism="Glycine max"
      /cultivar="Ripley"
      /db_xref="taxon:3847"
repeat_unit   1. .120
satellite     1. .240
            /note="STR120-B.1"
repeat_unit   121. .240
BASE COUNT      81 a 49 c 60 g 50 t
ORIGIN

Query Match      82.7%; Score 12.4; DB 8; Length 240;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
    |||||:|||||
Db 132 TCGAGTTTCGAGTT 119

RESULT 8
PUMPAL2A
LOCUS      PUMPAL2A      411 bp      DNA      PLN      21-JUL-1995
DEFINITION Petroselinum crispum phenylalanine ammonia-lyase (PAL-2) gene,
          promoter region and 5' end of cds.
ACCESSION L37355
VERSION   L37355.1 GI:567856
KEYWORDS  phenylalanine ammonia-lyase.
SOURCE    Petroselinum crispum DNA.
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.
REFERENCE 1 (bases 1 to 411)
AUTHORS  Logemann,E., Parniske,M. and Hahlbrock,K.
TITLE     Modes of expression and common structural features of the complete
          phenylalanine ammonia-lyase gene family in parsley
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (13), 5905-5909 (1995)
MEDLINE   95320184
FEATURES  Location/Qualifiers
  source
    1. .411
      /organism="Petroselinum crispum"
      /db_xref="taxon:4043"
      /note="(vector lambda EMBL4)"
promoter      1. .260
TATA_signal   230. .235
gene          409. .411
            /gene="PAL-2"
BASE COUNT      128 a 99 c 40 g 144 t
ORIGIN

Query Match      82.7%; Score 12.4; DB 8; Length 411;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
    |||||:|||||
Db 363 TCGAGTTTCGAGTT 376

RESULT 9
AX127755
LOCUS      AX127755      415 bp      DNA      PAT      15-MAY-2001
```

```

DEFINITION Sequence 8 from Patent WO0131042.
ACCESSION AX127755
VERSION AX127755.1 GI:14134402
SOURCE .
ORGANISM synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 415)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 8 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
Location/Qualifiers
1..415
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="5' border flanking region of elite event MS-B2"
misc_feature 1..234
/note="plant DNA"
misc_feature 235..415
/note="T-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN

Query Match 82.7%; Score 12.4; DB 6; Length 415;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
Db 2 TCGAGTTGGTGTT 15

RESULT 10
HSMCRP1/c HSMCRP1 636 bp DNA PRI 11-NOV-1998
LOCUS Homo sapiens MHC class I-related protein MRL precursor (MRL) gene,
signal peptide.
ACCESSION AF073484
VERSION AF073484.1 GI:3859892
KEYWORDS
SEGMENT 1 of 2
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 636)
AUTHORS Yamaguchi,H., Kuroiwa,Y. and Hashimoto,K.
TITLE Expanded genomic organization of conserved mammalian MHC class
I-related genes, human MRL and its murine ortholog
JOURNAL Biochem. Biophys. Res. Commun. 250 (3), 558-564 (1998)
MEDLINE 99003494
REFERENCE 2 (bases 1 to 636)
AUTHORS Hashimoto,K.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1998) Institute for Comprehensive Medical
Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan
FEATURES
Location/Qualifiers
1..636
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
/chromosome="1"
/map="1q25.3"
<1..45
/sig_peptide /gene="MRL"
BASE COUNT 160 a 140 c 138 g 198 t
ORIGIN

Query Match 82.7%; Score 12.4; DB 9; Length 636;
Best Local Similarity 71.4%; Pred. No. 2e+03;

```

```

ORGANISM Globodera rostochiensis
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
1 (bases 1 to 733) Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
AUTHORS Qin,L., Overmars,H., Helder,J., Popeijus,H., van der Voort,J.R.,
Groenink,W., van Koert,P., Schots,A., Bakker,J. and Smant,G.
TITLE An efficient cDNA-AFLP-based strategy for the identification of
putative pathogenicity factors from the potato cyst nematode
Globodera rostochiensis
JOURNAL Mol. Plant Microbe Interact. 13 (8), 830-836 (2000)
MEDLINE 20393248
REFERENCE 2 (bases 1 to 733)
AUTHORS Qin,L.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Qin L., Laboratory of Nematology, Wageningen
University, Binnenhaven 10, 6709 PD Wageningen, NETHERLANDS
FEATURES
source
1. .733
/organism="Globodera rostochiensis"
/isolate="Rol-Mierenbos"
/db_xref="taxon:31243"
/clones="A18"
/tissue_type="dorsal gland"
/dev_stage="infective second stage juvenile"
23. .73
23. .673
/note="ORF"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC21848.1"
/db_xref="GI:12227171"
/translation="MPTFLAAVCLLVADASPPKNTKLEKPSAGNAESTPGGLP
PQNRDPAHKKDLTLVDGLIVQNNKDKRWYVFAKQIPKDGIFVYVKILEKGNR
VOICLATQKPLDKPQWSEGTAYASAGYFWGKDKCTGTANGRPYKGPWFNG
DVICGVDFAHQIFTYKNGQRLITTELHVDHSAKLPCVSMFMPGKIEANFGPF"
BASE COUNT 210 a 168 c 171 g 184 t
ORIGIN

Query Match 82.7%; Score 12.4; DB 3; Length 733;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
||||:|:|:|
Db 105 TCGAGTFTGGTGT 92

RESULT 13
LOCUS GCU94186 905 bp mRNA PLN 20-DEC-1997
DEFINITION Colletotrichum gloeosporioides nitrogen starvation-induced
glutamine rich protein mRNA, complete cds.
ACCESSION U94186
VERSION U94186.1 GI:2707190
KEYWORDS
SOURCE Glomerella cingulata.
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
Glomerella.
REFERENCE 1 (bases 1 to 905)
AUTHORS Stephenson,S.-A., Maclean,D.J. and Manners,J.M.
TITLE Disruption of a novel pathogenicity gene of Colletotrichum
gloeosporioides results in a hypersensitive response in the host
Stylosanthes gulanensis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 905)
AUTHORS Stephenson,S.-A., Maclean,D.J. and Manners,J.M.
TITLE Direct Submission
JOURNAL Submitted (18-MAR-1997) CRCTPP, University of Queensland, St.Lucia,
Brisbane, Queensland 4072, Australia
FEATURES
Location/Qualifiers

Query Match 82.7%; Score 12.4; DB 8; Length 905;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
||||:|:|:|
Db 29 TCGAGTATCGAGT 16

RESULT 14
LOCUS AF049922 927 bp mRNA PLN 05-JUL-2000
DEFINITION Petunia x hybrida PGPS/D6 (PGPS/D6) mRNA, complete cds.
ACCESSION AF049922
VERSION AF049922.1 GI:4105781
KEYWORDS
SOURCE Petunia x hybrida.
ORGANISM Petunia x hybrida
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE 1 (bases 1 to 927)
AUTHORS Guyon,V.N., Astwood,J.D., Garner,E.C., Dunker,A.K. and Taylor,L.P.
TITLE Isolation and characterization of cDNAs expressed in the early
stages of flavonol-induced pollen germination in petunia
JOURNAL Plant Physiol. 123 (2), 699-710 (2000)
MEDLINE 20317212
PUBMED 10859200
REFERENCE 2 (bases 1 to 927)
AUTHORS Guyon,V., Astwood,J.D. and Taylor,L.P.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1998) Genetics and Cell Biology, Washington State
University, Science Hall, Pullman, WA 99164-4234, USA
FEATURES
source
1. .927
/organism="Petunia x hybrida"
/db_xref="taxon:4102"
/tissue_type="germinating petunia pollen treated with
kaempferol"
1. .927
/gene="PGPS/D6"
60. .707
/gene="PGPS/D6"
/note="ER lumen protein retaining receptor homolog;
putative HDEL receptor"
/codon_start=1
/product="PGPS/D6"
/protein_id="AAD02548.1"
/db_xref="GI:4105782"
/translation="MNIFRLAGDMTHLASVLVLLKTHITKSCAGSVLKTQELVALVF
VTRYLDITDFISLYNTMKLVFLGSSLSIVYMRHKIVRRSYDKDDOTFRHLFLVL
PCLLALVINEKFTFKEVMWTFISYLEAVAILPOLVLLQRTNIDNLTGQYIFLLGAY
RSFYILNNVRYFTPEPHVHWTWIAGLIQIOLLVADFFYYFQSKNNKLELPA"

```

BASE COUNT 273 a 173 c 176 g 305 t  
ORIGIN

## Query Match

Best Local Similarity 82.7%; Score 12.4; DB 8; Length 927;  
Matches 10; Conservative 71.4%; Pred. No. 2e+03;

Mismatches 4; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 694 TCGAGTTTCGAGTT 681

## RESULT 15

AF328994/c

LOCUS AF328994 942 bp DNA PLN 17-MAY-2001  
DEFINITION Arabidopsis lyrata putative S-glycoprotein (S) gene, S-13-4 allele, partial cds.

ACCESSION AF328994

VERSION AF328994.1 GI:13345390

## KEYWORDS

SOURCE

ORGANISM

Arabidopsis lyrata.

Arabidopsis lyrata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosoids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 942)

Schierup,M.H., Mable,B.K., Awadalla,P. and Charlesworth,D.

Identification and Characterization of a Polymorphic Receptor

Kinase Gene Linked to the Self-Incompatibility Locus of Arabidopsis

lyrata

Genetics 158 (1), 387-399 (2001)

21231627

11333247

2 (bases 1 to 942)

Schierup,M.H., Mable,B.K., Awadalla,P. and Charlesworth,D.

Direct Submission

Submitted (15-DEC-2000) Genetics and Ecology, University of Aarhus,

Ny Munkegade, Building 540, Aarhus 8000, Denmark

Location/Qualifiers

1..942

/organism="Arabidopsis lyrata"

/db\_xref="taxon:59689"

<1..>942

/gene="S"

/product="putative S-glycoprotein"

<1..>942

/gene="S"

/allele="S-13-4"

<1..>942

/gene="S"

/codon\_start=3

/product="putative S-glycoprotein"

/protein\_id="AAK1915.1"

/db\_xref="GI:13345391"

/translation="GNLVILGHSDKLWVSTNLTSGNVRSLVVAELLANGNFVRYSSK

DOGGFLWQSFYDTTLPQMKLGDRKTKLNRYLRSMKSSYDPSGGNFTYELETRGF

PEEFRLKTDIPHRSGPDGIRISGPEERQVDMVINYFTEDREVAAYTFLVTHNSIY

SLRTMSYSGYFRFTWTFPAGWQLWSAPMDLQCDLYPCGPYAYCDTTPPLCNCI

RGFNPMSEQNMWMDGTSGCVRRTPLSCRRDGLPMMKMLPTTTMTATVDRRISGKEC

KOKCLMDNCNTAYANADIKNGLGCVINTGELVDIRT"

260 a 182 c 249 g 251 t

BASE COUNT 260 a 182 c 249 g 251 t

ORIGIN

## Query Match

Best Local Similarity 82.7%; Score 12.4; DB 8; Length 942;  
Matches 10; Conservative 71.4%; Pred. No. 2e+03;

Mismatches 4; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|:|:|

Db 302 TCGAGTTTCGAGTT 289

Search completed: February 25, 2002, 18:00:08  
Job time: 18416 sec



---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:21 ; Search time 716.55 Seconds  
(without alignments)  
17.947 Million cell updates/sec

Title: US-09-698-903B-4  
Perfect score: 15  
Sequence: 1 ntgcstwtsgwgtt 15  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12.4	82.7	15	AAQ79175	Non-specific prime
2	12.4	82.7	15	AAA88455	PCR primer CA54 us
3	12.4	82.7	15	AAA48993	PCR primer MDB285
4	12.4	82.7	15	AAZ99438	PCR primer ARB1 fo
5	12.4	82.7	15	AAF90555	Degenerate primer
6	12.4	82.7	15	AAQ06993	PCR primer MDB285
7	12.4	82.7	15	AAF86681	Degenerate TAIL-PC
8	12.4	82.7	15	AAA89358	Arabidopsis DNA pr
9	12.4	82.7	15	AAAC86474	Maize Mac2 gene de
c 10	12.4	82.7	30	AAAC84167	Human pollinosis-a
11	12.4	82.7	304	AAH21085	Parallel detection

12	12.4	82.7	335	21	AAQ02398	Human secreted pro
13	12.4	82.7	336	20	AAQ51486	Human secreted pro
c 14	12.4	82.7	388	20	AAQ20974	Polynucleotide seq
15	12.4	82.7	415	22	AAQ06997	Right (5') border
16	12.4	82.7	633	22	AAH21072	Parallel detection
17	12.4	82.7	679	22	AAH99881	Human protein enco
18	12.4	82.7	802	21	AAQ64159	Human pollinosis-a
19	12.4	82.7	836	20	AAH81127	DNA sequence upstr
20	12.4	82.7	951	19	AAQ03315	Pyrolabus fumarius
21	12.4	82.7	1321	22	AAQ86489	Maize Mac2n-3 codi
22	12.4	82.7	2914	22	AAH17685	Human cDNA sequenc
23	12.4	82.7	3198	21	AAH48642	Arabidopsis thalia
24	12.4	82.7	4233	20	AAH87946	Candida tropicalis
25	12.4	82.7	4271	21	AAQ64160	Human pollinosis-a
26	12.4	82.7	4345	22	AAH99730	Human protein enco
c 27	12.4	82.7	5098	20	AAQ24984	E. coli MG1655 rin
c 28	12.4	82.7	49272	19	AAV35000	Mycobacteriophage
29	11.4	76.0	45	20	AAZ24665	Oligonucleotide fo
c 30	11.4	76.0	165	22	AAI55830	Probe #24516 used
31	11.4	76.0	249	19	AAV03550	DNA sequence that
32	11.4	76.0	438	21	AAQ52235	Arabidopsis thalia
33	11.4	76.0	462	22	AAH52999	S. epidermidis ope
34	11.4	76.0	482	21	AAQ38143	Zea mays DNA fragm
35	11.4	76.0	489	21	AAQ79427	Eucalyptus grandis
c 36	11.4	76.0	503	21	AAQ37052	Arabidopsis thalia
c 37	11.4	76.0	539	21	AAQ45066	Human secreted exp
c 38	11.4	76.0	574	21	AAQ09195	Fusarium venenatum
c 39	11.4	76.0	576	16	AAQ83845	Hepatitis C virus
40	11.4	76.0	576	17	AAQ16559	Hepatitis C virus
41	11.4	76.0	579	21	AAQ10353	Fusarium venenatum
42	11.4	76.0	581	22	AAH29686	Drosophila melanog
c 43	11.4	76.0	590	22	AAI42770	Probe #11456 used
44	11.4	76.0	606	22	AAH07816	Human cDNA clone (
45	11.4	76.0	619	21	AAQ48567	Arabidopsis thalia

ALIGNMENTS

RESULT 1  
AAQ79175  
ID AAQ79175 standard; DNA; 15 BP.  
XX  
AC AAQ79175;  
XX  
DT 22-JUN-1995 (first entry)  
XX  
DE Non-specific primer #1, for amplification of T-DNA.  
XX  
KW Polymerase chain reaction; PCR; amplify; primer; specific;  
KW non-specific; DNA polymerase; deoxyribonucleotide; cyclisation;  
KW ligation; ss.  
OS Synthetic.  
XX  
PN JP06253843-A.  
XX  
PD 13-SEP-1994.  
XX  
PF 03-MAR-1993; 93JP-0043006.  
XX  
PR 03-MAR-1993; 93JP-0043006.  
XX  
PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.  
XX (CHIK-) ZH CHIKYU KANKYO SANGYO GIJITSU KENKYU.  
XX  
DR WPI; 1994-328991/41.  
XX  
PT New DNA amplification method - without cyclisation of the DNA or  
XX ligation of oligonucleotide(s)  
PS Example 1; Page 10; 14pp; Japanese.  
XX

CC The sequences given in AAQ79172-74 are specific primers which were  
 CC used in conjunction with the non-specific primers given in AAQ79175-77  
 CC in the DNA amplification method of the invention. The method  
 CC comprises mixing a DNA sequence and a primer, DNA polymerase and  
 CC deoxyribonucleotides. The DNA is denatured allowing annealing of  
 CC the primers. The amplification cycle consists of one cycle at a  
 CC temperature to allow non-specific annealing of non-specific primers,  
 CC numerous cycles at a temp. to allow specific annealing of specific  
 CC primers and numerous cycles at a temp to allow non-specific primers  
 CC to hybridise specifically to complementary sequences. This method  
 CC allows amplification of neighbouring sequences with known sequences,  
 CC efficiently without cyclisation of the DNA or ligation of  
 CC oligonucleotides.  
 XX  
 SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 15; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 |||||  
 DB 2 tcgastwtsgwgtt 15

RESULT 2  
 AAA88455  
 ID AAA88455 standard; DNA; 15 BP.  
 XX  
 AC AAA88455;  
 XX  
 DT 09-JAN-2001 (first entry)  
 XX  
 DE PCR primer CA54 used to identify Arabidopsis 16713 gene.  
 XX  
 KW 16713 gene; essential gene; herbicide; screening; transgenic plant;  
 KW Arabidopsis thaliana; PCR primer; ss.  
 XX  
 OS Agrobacterium sp.  
 XX  
 PN WO200053782-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 03-MAR-2000; 2000WO-EP01884.  
 XX  
 PR 05-MAR-1999; 99US-0263385.  
 PR 10-MAY-1999; 99US-0309036.  
 PR 08-JUN-1999; 99US-0327662.  
 PR 29-SEP-1999; 99US-0408196.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Patton DA, Ashby CS, Thomas CR, Mcelver JA, Levin Jz;  
 PI Budziszewski GJ;  
 XX  
 XX WPI; 2000-594330/56.  
 XX  
 XX Novel genes isolated from Arabidopsis coding for proteins essential for  
 PT normal plant development, are useful to discover new herbicides and to  
 PT identify inhibitors that are potential herbicides -  
 XX  
 XX Example 4c; Page 53; 92pp; English.

CC This degenerate oligonucleotide, termed CA54, is a primer used in  
 CC the border rescue of the 16713 gene (see AAA88448) from Arabidopsis  
 CC thaliana embryo-lethal line #16713. DNA flanking the borders of  
 CC line #16713 was isolated using TAIL PCR. CA54 is 1 of 6 primers  
 CC (see AAA88451-56) used in combination with 2 sets of nested, T-DNA  
 CC specific primers for the right border (see AAA88457-59) as well as  
 CC for the left border (see AAA88460-62) of the T-DNA region of PCSA104.

CC The 16713 gene encodes a protein (see AAB19528) that is essential for  
 CC normal plant development. Chemicals that inhibit the function of  
 CC the 16713-encoded protein are likely to have detrimental effects on  
 CC plants and are potentially good herbicide candidates.

SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 |||||  
 DB 2 tcgastwtsgwgtt 15

RESULT 3  
 AAA48993  
 ID AAA48993 standard; DNA; 15 BP.  
 XX  
 AC AAA48993;  
 XX  
 DT 28-NOV-2000 (first entry)  
 XX  
 DE PCR primer MDB285 used to locate bar gene insertion in rice plant.  
 XX  
 KW Herbicide; phosphinothricin; glufosinate; bialaphos; bar;  
 KW phosphinothricin acetyl transferase; PAT; PCR primer; TAIL;  
 KW thermal asymmetric interlaced; rice; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200026345-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 03-NOV-1999; 99WO-US25667.  
 XX  
 PR 03-NOV-1998; 98US-0185244.  
 XX  
 PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Michiels F, Johnson K;  
 XX  
 DR WPI; 2000-365598/31.

XX New glufosinate tolerant rice plants with genomic DNA capable of  
 PT yielding at least 3 restriction fragments and having bar gene under  
 PT control of CaMV 35S promoter at specific location -  
 XX  
 XX Example 3; Page 25; 44pp; English.  
 XX  
 XX The present invention relates to a transgenic glufosinate tolerant rice  
 CC plant. The tolerant rice plants are generated by transformation with  
 CC the Streptomyces hygrosopicus bar gene. This gene encodes the enzyme  
 CC phosphinothricin acetyl transferase (PAT), which confers resistance to  
 CC the herbicidal compounds phosphinothricin (also called glufosinate) and  
 CC bialaphos. The 3' insertion site of the transgene was determined by  
 CC thermal asymmetric interlaced (TAIL) PCR. The present sequence is the  
 CC arbitrary degenerate PCR primer, MDB285, used for the TAIL PCR.  
 CC Glufosinate tolerant rice plants are useful as they give better yields  
 CC and growth, compared to untransformed plants, after application of a  
 CC glufosinate herbicide.

SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

Db 2 tcgastwtsgwgtt 15  
|||||

RESULT 4

AAZ99438  
ID AAZ99438 standard; DNA; 15 BP.

AC AAZ99438;  
XX

DT 03-JUL-2000 (first entry)  
XX

DE PCR primer ARB1 for the soybean AX5 promoter.  
XX

XX Gibberellic acid; copalyl diphosphate synthase; 3beta-hydroxylase;  
KW 2-oxidase; phytoene synthase; C-20 oxidase; 2beta,3beta-hydroxylase;  
KW seed germination; seedling growth; gibberellin biosynthetic pathway;  
KW transgenic plant; hypocotyl; epicotyl; AX5 promoter; PCR primer; ss.  
XX

OS Glycine max.  
XX

XX WO200009722-A2.  
XX

PN 24-FEB-2000.  
XX

PD 10-AUG-1999; 99WO-US18066.  
XX

PF 10-AUG-1999; 98US-0096111.  
XX

PR 07-JUN-1999; 99US-0137977.  
XX

XX (MONS ) MONSANTO CO.  
PA

XX Brown SM, Elich TD, Heck GR, Kishore GM, Logusch EW, Logusch SJ;  
PI Piller KJ, Rao S, Ream JE;  
XX WPI; 2000-224351/19.

XX Obtaining transgenic plant useful for controlling seed germination and  
PT seedling growth comprises transgene comprising a sequence expressing  
PT altered levels of an essential hormone -  
XX

PS Example 7; Page 98; 267pp; English.  
XX

XX PCR primers AAZ99438-42 were used to amplify the soybean AX5 promoter.  
CC The amplified sequence is used in the method of the invention. The  
CC specification describes methods for the inhibition and control of  
CC gibberellic acid levels. Gibberellic acid levels may be inhibited or  
CC controlled by use of a chimeric expression construct expressing a RNA  
CC or protein which suppresses the gibberellin biosynthetic pathway  
CC sequence, diverts substrate from the pathway, or degrades pathway  
CC substrates or products. The methods uses copalyl diphosphate synthase,  
CC 3beta-hydroxylase, 2-oxidase, phytoene synthase, C-20 oxidase, and  
CC a 2beta,3beta-hydroxylase polynucleotides to achieve this. The method  
CC is used to control seed germination and seedling growth especially to  
CC regulate gene products of gibberellin biosynthetic pathway and  
CC restoration of normal seed germination, in transgenic plants. The  
CC plants produced are gibberellin deficient, and have shortened hypocotyl  
CC and/or epicotyl phenotypes compared to normal plants.  
XX

SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 21; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||

Db 2 tcgastwtsgwgtt 15  
|||||

RESULT 5

AAF90555

ID AAF90555 standard; DNA; 15 BP.

XX AAF90555;  
AC

XX 22-AUG-2001 (first entry)  
DT

XX Degenerate primer CA54 used to prime Arabidopsis genomic DNA.  
DE

XX ET1158 gene; GT6839 gene; ET5262 gene; herbicide; screening;  
KW herbicide tolerance; transgenic plant; crop protection; PCR primer;  
KW ss.  
XX

XX Synthetic.  
OS

XX WO200144277-A2.  
PN

XX 21-JUN-2001.  
XX

XX 14-DEC-2000; 2000WO-EP12748.  
PF

XX 16-DEC-1999; 99US-0465040.  
PR

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX

XX Wegrich Glover L, Budziszewski GJ, Levin JZ, Zhou Q;  
PI WPI; 2001-398122/42.

XX New herbicide target genes encoding proteins having ET1158, GT6839 or  
PT ET5262 activity, for identifying an inhibitor of protein activity -  
PT

XX Example 1; Page 39; 67pp; English.  
PS

XX The present sequence is that of degenerate primer CA54. Arbitrary  
CC degenerate primers LWAD1, CA51, CA52, CA53, CA54 and CA55 (see  
CC AAF90551-56) were used to prime Arabidopsis thaliana genomic DNA  
CC flanking a Ds transposon insertion. The degenerate primers were  
CC used in combination with 2 sets of 3, nested, transposon-specific  
CC primers (see AAF90557-62) that were homologous to regions of the  
CC Ds elements which lie at the outermost ends of the transposons. A  
CC series of low- and high-stringency PCR amplifications were  
CC performed using the TAIL-PCR protocol. DNA fragments were produced  
CC which corresponded to the genomic DNA that was directly adjacent to  
CC the transposon insertion. Sequence analysis of PCR products from  
CC tagged seedling lethal lines ET1158, GT6839 and ET5262 identified 3  
CC novel genes (see AAF90548-50) each of which was essential for  
CC Arabidopsis seedling growth and development. The essentiality of  
CC the genes provides a means of discovering new herbicides. Screening  
CC assays for identifying inhibitors that are potential herbicides are  
CC provided. The invention is also applied to the development of  
CC herbicide tolerant plants, and plant tissues, seeds and cells.  
XX

SQ Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||

Db 2 tcgastwtsgwgtt 15  
|||||

RESULT 6

AAD06993

ID AAD06993 standard; DNA; 15 BP.  
XX

XX AAD06993;  
AC

XX 06-AUG-2001 (first entry)  
DT

XX PCR primer MDB285 to generate the flanking region of elite event MS-B2.  
DE

XX MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; ss.  
 XX Unidentified.  
 XX WO200131042-A2.  
 XX 03-MAY-2001.  
 XX 26-OCT-2000; 2000WO-EPI0680.  
 XX 29-OCT-1999; 99US-0430497.  
 XX (AVET ) AVENTIS CROPS SCIENCE NV.  
 XX Weston B, De Beuckeleer M;  
 XX WPI; 2001-300517/31.  
 XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -  
 XX Example 3; Page 28; 53pp; English.  
 XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is degenerate PCR primer M8285 used to generate  
 CC right (5') and left (3') border flanking region of elite event MS-B2.  
 XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
 SQ  
 Query Match 82.7%; Score 12.4; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 tcgastwtsgwgtt 15  
 Db | | | | | | | | | | | | | | | |  
 2 tcgastwtsgwgtt 15  
 RESULT 7  
 AAF86681  
 ID AAF86681 standard; DNA; 15 BP.  
 XX AAF86681;  
 XX 25-JUL-2001 (first entry)  
 XX Degenerate TAIL-PCR primer AD1, used for mapping insertion sites.  
 XX Transgenic plant; cross-fertilisation; cross-breeding;  
 KW modified gene activity; expression; timing; tissue specificity;  
 KW modified nucleic acid structure; nucleic acid effector molecule;  
 KW activator; methylation enzyme; repressor; gyrase;  
 KW kinase; topoisomerase; class I restriction enzyme; recognition sequence;  
 KW phenotype selection; Gramineae; Solanaceae; Leguminosae;  
 KW insertion site mapping; T-DNA; Ds element; rice; TAIL PCR primer;  
 KW thermal asymmetric interlaced PCR; ss.  
 XX Synthetic.  
 OS  
 XX WO200121781-A2.  
 PN  
 XX 29-MAR-2001.  
 PD  
 XX

PF 20-SEP-2000; 2000WO-US25778.  
 XX 20-SEP-1999; 99US-0155036.  
 XX (MOLE-) CENT APPL MOLECULAR BIOLOGY INT AGRIC.  
 XX Jefferson RA, Killian A, Nottenburg C, Keese PK, Mayer J;  
 PI Stachel SE;  
 XX WPI; 2001-335475/35.  
 XX Generating transgenic plants with modified genetic activity or nucleic  
 PT acid structure comprises crossing transgenic plant with gene for  
 PT nucleic acid effector with transgenic plant with binding site for  
 PT effector -  
 XX Example 3; Page 27; 42pp; English.  
 XX The invention relates to a method of generating transgenic plants  
 CC that have modified gene activity or nucleic acid structure. The method  
 CC involves the cross-fertilisation of two transgenic plant lines to produce  
 CC seed, which is then grown to produce a plant. One of the parental plant  
 CC lines contains a gene encoding a non-native site-specific nucleic acid  
 CC effector molecule under the control of a minimal promoter and operably  
 CC linked to an endogenous enhancer element, and the other contains a  
 CC recognition sequence to which the nucleic acid effector molecule can  
 CC specifically bind. The nucleic acid effector molecule may be a  
 CC transcriptional activator, methylation enzyme, repressor, gyrase,  
 CC kinase, topoisomerase, class I restriction enzyme, or single or double  
 CC stranded DNA or RNA. In another embodiment of the invention, three  
 CC transgenic plant lines are crossed, in which a transcriptional activator  
 CC gene, the minimal promoter and the recognition site are contained within  
 CC the three separate plant lines. In the transgenic cross-breed plant,  
 CC binding of the nucleic acid effector molecule to the recognition sequence  
 CC modifies gene expression, the timing of gene expression, the tissue  
 CC specificity of gene expression, or nucleic acid structure. Such plants  
 CC displaying a desired phenotype can then be selected. The method of the  
 CC invention can be used to generate transgenic plants, including Gramineae  
 CC (e.g., wheat, maize, rice, barley, rye, oats, sugarcane), Solanaceae  
 CC (e.g., potato, tomato, peppers), and Leguminosae (e.g., beans, soybean,  
 CC lentil, chickpea and peanut). Sequences AAF86681-AAF86683 represent  
 CC degenerate primers used in TAIL-PCR (thermal asymmetric interlaced PCR)  
 CC for physical mapping of T-DNA and Ds element insertion elements in  
 CC transgenic rice plants.  
 XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
 SQ  
 Query Match 82.7%; Score 12.4; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 tcgastwtsgwgtt 15  
 Db | | | | | | | | | | | | | | | |  
 2 tcgastwtsgwgtt 15  
 RESULT 8  
 AAA89358  
 ID AAA89358 standard; DNA; 15 BP.  
 XX AAA89358;  
 AC AAA89358;  
 XX 23-APR-2001 (first entry)  
 DT Arabidopsis DNA primer AD1.  
 XX Arabidopsis silencing; PCR primer; ss.  
 XX Gene silencing; PCR primer; ss.  
 XX Arabidopsis thaliana.  
 OS  
 XX WO200100801-A2.  
 PN  
 XX

PD 04-JAN-2001.  
 XX  
 XX  
 PF 21-JUN-2000; 2000WO-EP05761.  
 XX  
 PR 23-JUN-1999; 99GB-0014623.  
 XX  
 XX (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 XX  
 XX  
 PI Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;  
 XX  
 XX WPI; 2001-137952/14.  
 DR  
 XX  
 XX Novel gene encoding a protein that controls gene silencing, in  
 PT particular silencing of plant genes -  
 XX  
 XX Example 4; Page 11; 48pp; English.  
 PS  
 XX This is the sequence of primer Ad1, which is 1 of 7 degenerate  
 CC primers (see AA89358-64) expected to bind to Arabidopsis thaliana  
 CC DNA flanking the site of a T-DNA insert. These primers were used  
 CC with 3 specific, nested primers (see AA89355-57), located close to  
 CC the right border of T-DNA, in the cloning of the silencing gene  
 CC (see AA89353) of Arabidopsis thaliana. Genomic DNA from the plant  
 CC containing only T-DNA co-segregating with a hygromycin resistant  
 CC mutant phenotype was isolated. Gene silencing is useful as a  
 CC molecular tool for regulating gene expression.  
 XX  
 XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
 SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 tcgastwtsgwgtt 15  
 Db 2 tcgastwtsgwgtt 15  
 |||||||  
 RESULT 9  
 AAC86474  
 ID AAC86474 standard; DNA; 15 BP.  
 XX  
 AC AAC86474;  
 XX  
 DT 01-MAR-2001 (first entry)  
 XX  
 XX Maize Mac2 gene degenerate PCR primer Ad1.  
 DE  
 XX Tapetum-specific promoter; artificial male sterility; agriculture;  
 KW crop yield; disease resistance; AMS; maize; MAC2; PCR primer; ss.  
 KW  
 XX  
 OS Zea mays.  
 XX  
 FN WO200068403-A2.  
 XX  
 PD 16-NOV-2000.  
 XX  
 XX 10-MAY-2000; 2000WO-GB01789.  
 PF  
 XX  
 PR 10-MAY-1999; 99GB-0010796.  
 XX  
 PA (BT0G-) BIOGEWA UK LTD.  
 XX  
 XX Paul W, Scott RJ, Hird D, Hodge R;  
 PI  
 XX WPI; 2001-016099/02.  
 DR  
 XX Novel nucleic acid sequences encoding tapetum specific promoters, pMAC2  
 PT and pMAC20, or promoter sequences that control the expression of coding  
 PT sequence homologous to the promoters, for artificial male sterility -  
 XX

FS Example 2; Page 16; 45pp; English.  
 XX  
 XX The present invention provides tapetum-specific promoters which can be  
 CC used to create artificial male sterility (AMS) systems in plants. These  
 CC are useful in agriculture as they enable the production of plants with  
 CC increased yield and greater disease resistance. They also result in  
 CC uniform offspring, which have the same germination time, height of  
 CC growth, susceptibility to disease, flowering time etc.  
 XX  
 XX Sequence 15 BP; 1 A; 1 C; 3 G; 5 T; 5 other;  
 SQ

Query Match 82.7%; Score 12.4; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 tcgastwtsgwgtt 15  
 Db 2 tcgastwtsgwgtt 15  
 |||||||  
 RESULT 10  
 AAC64167/c  
 ID AAC64167 standard; DNA; 30 BP.  
 XX  
 AC AAC64167;  
 XX  
 DT 21-FEB-2001 (first entry)  
 XX  
 XX Human pollinosis-associated gene 581 hybridisation probe, SEQ ID NO:8.  
 DE  
 XX Human; pollinosis-associated gene 581; IgE; immunoglobulin E;  
 KW cedar pollen allergy; T-cell; reduced expression; detection;  
 KW diagnosis; drug screening; allergic disease; expression analysis;  
 KW hybridisation probe; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200065048-A1.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 26-APR-2000; 2000WO-JP02732.  
 XX  
 PR 27-APR-1999; 99JP-0120492.  
 XX  
 PA (GENO-) GENOX RES INC.  
 XX  
 PI Nagasu T, Sugita Y, Kashiwabara T, Oshida T, Obayashi M, Gunji S;  
 PI Obayashi I, Imai Y, Yoshida N, Ogawa K, Matsui K;  
 XX  
 DR WPI; 2000-687341/67.  
 XX  
 XX Pollenosis-associated gene 581 undergoing significantly low expression  
 PT in subjects with high cedar pollen-specific IgE levels, useful in  
 PT diagnosis of allergic diseases and screening drug candidates -  
 XX  
 XX Example 8; Page 24; 69pp; Japanese.  
 FS  
 XX The invention relates to the human pollinosis-associated gene 581 which  
 CC exhibits significantly reduced expression in the T-cells of individuals  
 CC with high cedar pollen-specific IgE (immunoglobulin E) levels. The gene  
 CC was isolated from T-cells from individuals allergic to cedar pollen using  
 CC the differential display method. The invention also relates to cedar pollen  
 CC constructs and host cells comprising pollinosis-associated gene 581;  
 CC to the protein encoded by pollinosis-associated gene 581; to expression  
 CC nucleic acids; pollinosis-associated gene 581 primers and probes;  
 CC antibodies against the protein encoded by the gene; methods of detection  
 CC of pollinosis-associated gene 581 nucleic acids; and a method of  
 CC diagnosis of allergic diseases via the detection of pollinosis-associated  
 CC gene 581 nucleic acids. The invention additionally encompasses methods of  
 CC screening drug candidates for the treatment of allergic disease by  
 CC measuring the expression of pollinosis-associated gene 581 in pollen

CC antigen-stimulated T-cells in the presence of a test compound relative  
 CC to a control. Pollinosis-associated gene 581 is useful in the diagnosis  
 CC of allergic diseases and in the screening of drug candidates for the  
 CC treatment of such diseases. The present sequence represents a  
 CC hybridisation probe used in human pollinosis-associated gene 581  
 CC expression analysis.

XX Sequence 30 BP; 9 A; 9 C; 4 G; 8 T; 0 other;

Query Match 82.7%; Score 12.4; DB 21; Length 30;  
 Best Local Similarity 71.4%; Pred. No. 3.2e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 Db 30 TCGAGTATGGTGT 17  
 ||||:|:|:|

## RESULT 11

AAH21085  
 ID AAH21085 standard; DNA; 304 BP.

AC AAH21085;

DT 03-SEP-2001 (first entry)

XX Parallel detection of methylated genomic DNA associated DNA SEQ ID 31.

DE DNA methylation; parallel detection; 5-unmethylated cytosine; CpG;  
 KW CpnPG; amplification; transcription regulation; genetic imprinting;  
 KW tumorigenesis; primer; ss.

OS Unidentified.

WO200142493-A2.

PN 14-JUN-2001.

XX 06-DEC-2000; 2000WO-DE04381.

XX 06-DEC-1999; 99DE-1059691.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C;

XX WPI: 2001-381705/40.

DR Parallel detection of the methylation pattern of many genomic DNA  
 XX regions, useful for detecting aberrant methylation, includes multiple  
 PT amplification of chemically modified DNA -

PS Disclosure; Page 63; 63pp; German.

XX This invention describes a novel method for the parallel detection of the  
 CC methylation status of genomic DNA (1) which involves a (1) sample being  
 CC treated chemically to convert 5-unmethylated cytosine to uracil,  
 CC thymidine or some other base having hybridization behavior different from  
 CC that of C, then amplifying simultaneously at least 10 different fragments  
 CC (of fewer than 2 kb) using synthetic oligonucleotide (ON) primers. These  
 CC primers are based on regulatory, transcribed and/or translated segments  
 CC present in the sample after chemical treatment. The sequence context of  
 CC all, or some, of the CpG and CpNpG motifs in the amplified products is  
 CC then determined. The method is used to detect aberrant methylation  
 CC patterns in the genome, these are implicated in regulation of  
 CC transcription, genetic imprinting and tumorigenesis. Many target regions  
 CC in the genome can be analyzed simultaneously and it is not essential to  
 CC know the sequence context of all targeted regions. Primers may be  
 CC designed for preferential amplification of particular segments of  
 CC interest (e.g. promoters and exons).

XX Sequence 304 BP; 37 A; 35 C; 131 G; 101 T; 0 other;

Query Match 82.7%; Score 12.4; DB 22; Length 304;  
 Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 Db 112 tcgagtttggagtt 125  
 ||||:|:|:|

## RESULT 12

AAC02398  
 ID AAC02398 standard; cDNA; 335 BP.

XX AAC02398;

DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 2396.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

PA (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX P-PSDB; AAC02392.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PS Claim 1; SEQ ID 2396; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.

XX Sequence 335 BP; 71 A; 75 C; 89 G; 89 T; 11 other;

Query Match 82.7%; Score 12.4; DB 21; Length 335;  
 Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
 Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
 Db 194 tcgagtttggagtt 207  
 ||||:|:|:|



```

RESULT 13
AAX51486
ID AAX51486 standard; cDNA; 336 BP.
XX
AC AAX51486;
XX
DT 21-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST SEQ ID NO:65.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN WO9906549-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01231.
XX
PR 01-AUG-1997; 97US-0905279.
XX
PA (GEST ) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153779/13.
XX
DR P-PSDB; AAY12708.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX
PS Claim 1; Page 192; 522pp; English.
XX
CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12681 to
CC AAY12913, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 336 BP; 72 A; 75 C; 89 G; 89 T; 11 other;

Query Match 82.7%; Score 12.4; DB 20; Length 336;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
Db 195 tcgagtttgaggtt 208
||||:|:|:|:|

RESULT 14
AAX20974/c
ID AAX20974 standard; DNA; 388 BP.

```

```

XX AAX20974;
XX AC
XX DT 05-MAY-1999 (first entry)
XX DE
XX DE Polynucleotide sequence from the genome of Treponema pallidum.
XX KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX KW enzyme production; ds.
XX OS Treponema pallidum.
XX PN WO9859034-A2.
XX PD 30-DEC-1998.
XX PF 23-JUN-1998; 98WO-US13041.
XX PR 24-JUN-1997; 97US-0050667.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Fraser CM;
XX DR WPI; 1999-081273/07.
XX PT New isolated Treponema pallidum nucleic acids - used to develop
XX PT products for the detection, diagnosis, characterisation, prevention
XX PT and therapy of T. pallidum infections, particularly syphilis
XX PS Claim 1; Page 998-999; 1150pp; English.
XX CC AAX20500-21243 represent polynucleotide sequences from the genome of
XX CC Treponema pallidum. The sequences can be used for detection
XX CC diagnosis, characterisation, prevention and therapy for T. pallidum
XX CC infections, particularly syphilis. They can also be used for detecting
XX CC diseases related to Borrelia infections in animals, and for the
XX CC production of biosynthetic products such as enzymes.
XX SQ Sequence 388 BP; 121 A; 68 C; 77 G; 117 T; 5 other;

Query Match 82.7%; Score 12.4; DB 20; Length 388;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15
Db 323 TCGAGTTTGGTGT 310
||||:|:|:|:|

RESULT 15
AAD06997
ID AAD06997 standard; DNA; 415 BP.
XX AC AAD06997;
XX DT 06-AUG-2001 (first entry)
XX DE
XX DE Right (5') border flanking region of elite event MS-B2.
XX KW MS-B2 elite event; transgenic Brassica plant; transformation event;
XX KW male-sterility gene; ds.
XX OS Chimeric - Agrobacterium sp.
XX OS Chimeric - Brassica sp.
XX PH Key Location/Qualifiers
XX FT misc_feature 1..234
XX FT /tag= a
XX FT /note= "Corresponds to plant DNA"
XX FT misc_feature 235..415
XX FT /tag= b

```

/note= "Corresponds to T-DNA"

FT WO200131042-A2.  
XX  
XX  
XX PD 03-MAY-2001.  
XX  
XX PF 26-OCT-2000; 2000WO-EF10680.  
XX  
XX PR 29-OCT-1999; 99US-0430497.  
XX  
XX PA (AVET ) AVENTIS CROPS SCIENCE NV.  
XX  
XX PI Weston B, De Beuckeleer M;  
XX  
XX DR WPI; 2001-300517/31.  
XX  
XX PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX  
XX PS Claim 11; Page 51; 53pp; English.  
XX  
XX CC The present invention relates to a transgenic Brassica plant or its  
CC seed, cells or tissues, characterised by harbouring a specific  
CC transformation event, particularly by the presence of a male-sterility  
CC gene, at a specific location in the Brassica genome. Transgenic  
CC Brassica plant is useful for producing a hybrid seed by crossing the  
CC transgenic plant with a male-fertile Brassica plant and harvesting the  
CC hybrid seed from the transgenic Brassica plant.  
CC The present sequence is right (5') border flanking region of elite event  
CC MS-B2.  
XX  
XX SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 82.7%; Score 12.4; DB 22; Length 415;  
Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
| | | | : | : | | |  
Db 2 tcgagtttggtt 15

Search completed: February 25, 2002, 18:17:22  
Job time: 16680 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:11 ; Search time 301.6 Seconds  
(without alignments)  
11.264 Million cell updates/sec

Title: US-09-698-903B-4  
Perfect score: 15  
Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCRU5\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12.4	82.7	15	4	US-09-503-391-13
c 2	12.4	82.7	49272	1	US-08-614-770A-1
3	11.4	76.0	576	1	US-08-086-428B-1
4	11.4	76.0	576	2	US-08-468-570-1
5	11.4	76.0	576	2	US-08-290-665A-1
6	11.4	76.0	576	5	PCT-US95-10398-1
7	11.4	76.0	1268	4	US-09-046-894-50
c 8	11.4	76.0	2220	5	PCT-US95-13749-2
c 9	11.4	76.0	2341	3	US-09-187-049-11
10	11.4	76.0	2493	3	US-08-804-439A-11
11	11.4	76.0	2493	3	US-08-720-229-11
12	11.4	76.0	2652	2	US-08-953-492-1
13	11.4	76.0	2873	1	US-08-149-695-1
14	11.4	76.0	2873	1	US-08-377-228-1
15	11.4	76.0	3139	2	US-08-693-457-1
16	11.4	76.0	3139	4	US-09-265-731-1
17	11.4	76.0	12311	4	US-08-750-717-1
c 18	11.2	74.7	1567	1	US-08-181-271A-100
c 19	11.2	74.7	1567	1	US-08-449-315-100
c 20	11.2	74.7	1567	1	US-08-444-803-100
c 21	11.2	74.7	1567	1	US-08-449-043-100
c 22	11.2	74.7	1567	1	US-08-456-265A-100
c 23	11.2	74.7	1567	1	US-08-455-416-100
c 24	11.2	74.7	1567	1	US-08-455-244-100
c 25	11.2	74.7	1567	1	US-08-454-876-100
c 26	11.2	74.7	1567	2	US-08-457-364-100
c 27	11.2	74.7	1567	2	US-08-456-362-100

c 28	11.2	74.7	1567	2	US-08-456-240-100	Sequence 100, App
c 29	11.2	74.7	1567	2	US-08-455-736-100	Sequence 100, App
c 30	11.2	74.7	1567	2	US-08-971-217-100	Sequence 100, App
c 31	11.2	74.7	1567	4	US-09-350-600-100	Sequence 100, App
32	11.2	74.7	2068	1	US-08-309-341-1	Sequence 1, Appli
33	11.2	74.7	2068	1	US-08-608-267-1	Sequence 1, Appli
34	11.2	74.7	2068	1	US-08-608-452-1	Sequence 1, Appli
35	11.2	74.7	2068	1	US-08-608-224-1	Sequence 1, Appli
36	11.2	74.7	2068	2	US-08-967-149-1	Sequence 1, Appli
c 37	11.2	74.7	3117	1	US-08-172-331B-3	Sequence 3, Appli
c 38	11.2	74.7	3280	1	US-08-259-000-4	Sequence 4, Appli
c 39	11.2	74.7	3280	1	US-08-729-767-6	Sequence 6, Appli
c 40	11.2	74.7	4643	2	US-08-605-106-6	Sequence 6, Appli
41	11.2	74.7	7400	1	US-07-674-852-1	Sequence 1, Appli
42	11.2	74.7	7400	3	US-08-473-185-1	Sequence 1, Appli
43	11.2	74.7	7400	4	US-08-171-387-3	Sequence 3, Appli
c 44	11.2	74.7	12494	4	US-09-935-312-13	Sequence 13, Appli
c 45	11.2	74.7	12494	4	US-08-848-760B-33	Sequence 33, Appli

ALIGNMENTS

RESULT 1  
US-09-503-391-13  
; Sequence 13, Application US/09503391  
; Patent No. 6300091  
; GENERAL INFORMATION:  
; APPLICANT: Patton, David A.  
; APPLICANT: Ashby, Carl S.  
; APPLICANT: Thomas, Carla R.  
; APPLICANT: McElver, John A.  
; APPLICANT: Budziszewski, Gregory J.  
; APPLICANT: Levin, Joshua Z.  
; TITLE OF INVENTION: Herbicide Target Genes and Methods  
; FILE REFERENCE: PB/5-30852A  
; CURRENT APPLICATION NUMBER: US/09/503,391  
; CURRENT FILING DATE: 2000-02-14  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 15  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: oligonucleotide  
US-09-503-391-13

Query Match 82.7%; Score 12.4; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
|||||  
Db 2 tcgastwtsgwgtt 15

RESULT 2  
US-08-614-770A-1/c  
; Sequence 1, Application US/08614770A  
; Patent No. 5773267  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL  
; TITLE OF INVENTION: D29 SHUTTLE GRAHMS AND USES THEREOF  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN  
; STREET: 90 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.

ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE  
MEDIUM TYPE: DISKETTE  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/614,770A  
FILING DATE: MARCH 7, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 96700/402  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49272  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: OLIGONUCLEOTIDE  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: MYCOBACTERIOPHAGE  
INDIVIDUAL ISOLATE: D29  
US-08-614-770A-1

Query Match 82.7%; Score 12.4; DB 1; Length 49272;  
Best Local Similarity 71.4%; Pred. No. 1e+02; Indels 0; Gaps 0;  
Matches 10; Conservative 4; Mismatches 0;

QY 2 tcgastwtsgwgtt 15  
||||:|:|:|

Db 29087 TCGAGTTCGTGT 29074

RESULT 3  
US-08-086-428B-1  
; Sequence 1, Application US/08086428B  
; Patent No. 5514539  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R. H. AND  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 159  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/086,428B  
; FILING DATE: 29-JUN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4070  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 576 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; ORIGINAL SOURCE:  
; ORGANISM: homosapiens  
; INDIVIDUAL ISOLATE: DK7  
; US-08-086-428B-1

Query Match 76.0%; Score 11.4; DB 1; Length 576;  
Best Local Similarity 69.2%; Pred. No. 3.4e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgt 14  
||||:|:|:|

Db 55 TCGAGTATCGTGT 67

RESULT 4  
US-08-468-570-1  
; Sequence 1, Application US/08468570  
; Patent No. 5871962  
; GENERAL INFORMATION:  
; APPLICANT: BUKH, J., MILLER, R. H. AND  
; APPLICANT: PURCELL, R. H.  
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE  
; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE  
; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN  
; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES  
; NUMBER OF SEQUENCES: 159  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,570  
; FILING DATE: 6-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/086,428  
; FILING DATE: 29-JUN-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RICHARD W. BORK  
; REGISTRATION NUMBER: 36,459  
; REFERENCE/DOCKET NUMBER: 2026-4070US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 576 base pairs  
; TYPE: nucleic acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DK7
US-08-468-570-1

Query Match          76.0%; Score 11.4; DB 2; Length 576;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14
Db 55 TCGAGTATCGTGT 67

RESULT 5
US-08-290-665A-1
; Sequence 1, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290.665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DK7
US-08-290-665A-1

Query Match          76.0%; Score 11.4; DB 2; Length 576;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14
Db 55 TCGAGTATCGTGT 67

US-08-290-665A-1

; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DK7
US-08-468-570-1

Query Match          76.0%; Score 11.4; DB 2; Length 576;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14
Db 55 TCGAGTATCGTGT 67

RESULT 6
PCT-US95-10398-1
; Sequence 1, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DK7
PCT-US95-10398-1

Query Match          76.0%; Score 11.4; DB 5; Length 576;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14
Db 55 TCGAGTATCGTGT 67

RESULT 7
US-09-046-894-30
; Sequence 30, Application US/09046894
; Patent No. 6190857
; GENERAL INFORMATION:
; APPLICANT: Ralph, David
; APPLICANT: An, Gang
; APPLICANT: O'Hara, Mark S.
; APPLICANT: O'Hara, Mark S.
```

APPLICANT: Veltri, Robert  
TITLE OF INVENTION: DIAGNOSIS OF DISEASE STATE USING mRNA  
TITLE OF INVENTION: PROFILES IN PERIPHERAL LEUKOCYTES  
NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,894  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/041,576  
FILING DATE: 24-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakashima, Richard A.  
REGISTRATION NUMBER: P-42,023  
REFERENCE/DOCKET NUMBER: UROC:014  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1268 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-046-894-30

Query Match 76.0%; Score 11.4; DB 4; Length 1268;  
Best Local Similarity 69.2%; Pred. No. 3.5e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 cgaestwtsgwgtt 15  
||||:|:|:|

Db 850 CGAGTTTCGGT 862

RESULT 8  
PCT-US95-13749-2/c  
Sequence 2, Application PC/TUS9513749  
GENERAL INFORMATION:  
APPLICANT: Amgen Inc.  
TITLE OF INVENTION: COMPOSITIONS FOR INCREASED  
TITLE OF INVENTION: BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavenland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13749  
FILING DATE:  
CLASSIFICATION:  
INFORMATION FOR SEQ ID NO: 2:  
.

SEQUENCE CHARACTERISTICS:  
LENGTH: 2220 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 536..1024  
PCT-US95-13749-2

Query Match 76.0%; Score 11.4; DB 5; Length 2220;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgt 14  
||||:|:|:|

Db 975 TCGAGTTTGGAGT 963

RESULT 9  
US-09-187-049-11/c  
Sequence 11, Application US/09187049  
Patent No. 6117666  
GENERAL INFORMATION:  
APPLICANT: Lamppa, Gayle K.  
TITLE OF INVENTION: PLASTID PROTEOLYTIC PROCESSING ENZYME  
TITLE OF INVENTION: THAT CLEAVES PRECURSOR POLYPEPTIDES  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BRINKS HOFER GILSON & LIONE  
STREET: P.O. Box 10395  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/187,049  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/695,177  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Martin, Alice O.  
REGISTRATION NUMBER: 35,601  
REFERENCE/DOCKET NUMBER: 7814/16  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312 321-4200  
TELEFAX: 312 321-4299  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2341 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Arabidopsis CPE  
US-09-187-049-11

Query Match 76.0%; Score 11.4; DB 3; Length 2341;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwt 14  
 |||||:|:|:|  
 Db 599 TCGACTATGGAGT 587

RESULT 10  
 US-08-804-439A-11  
 ; Sequence 11, Application US/08804439A  
 ; Patent No. 6015565  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, Timothy M.  
 ; APPLICANT: Bosch, Marnix L.  
 ; APPLICANT: Strand, Kurt  
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RHV/KSHV  
 ; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
 ; NUMBER OF SEQUENCES: 113  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Ste 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/804,439A  
 ; FILING DATE: February 21, 1997  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 09176/004001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 678-5070  
 ; TELEFAX: (619) 678-5099  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2493 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-804-439A-11

Query Match 76.0%; Score 11.4; DB 3; Length 2493;  
 Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 tcgastwtsgwt 14  
 |||||:|:|:|  
 Db 1369 TCGAGTATCGTGT 1381

RESULT 11  
 US-08-720-229-11  
 ; Sequence 11, Application US/08720229  
 ; Patent No. 6022542  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, Timothy M.  
 ; APPLICANT: Bosch, Marnix L.  
 ; APPLICANT: Strand, Kurt  
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RHV/KSHV  
 ; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
 ; NUMBER OF SEQUENCES: 100  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morrison & Foerster  
 ; STREET: 755 Page Mill Road

; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304-1018  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/720,229  
 ; FILING DATE: 26-SEP-1996  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Schiff, J. Michael  
 ; REGISTRATION NUMBER: 40,253  
 ; REFERENCE/DOCKET NUMBER: 29938-20002.00  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 813-5600  
 ; TELEFAX: (415) 494-0792  
 ; TELEX: 706141  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2493 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-720-229-11

Query Match 76.0%; Score 11.4; DB 3; Length 2493;  
 Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 tcgastwtsgwt 14  
 |||||:|:|:|  
 Db 1369 TCGAGTATCGTGT 1381

RESULT 12  
 US-08-953-492-1  
 ; Sequence 1, Application US/08953492  
 ; Patent No. 5849555  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brown, James  
 ; APPLICANT: Jaworski, Deborah  
 ; APPLICANT: Lawlor, Elizabeth  
 ; APPLICANT: Wang, Min  
 ; TITLE OF INVENTION: NOVEL vals  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SmithKline Beecham Corporation  
 ; STREET: 709 Swedeland Road  
 ; CITY: King of Prussia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19406-0939  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/953,492  
 ; FILING DATE: 17-OCT-1997  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/844,064  
 ; FILING DATE: 18-APR-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 9607991.8  
 ; FILING DATE: 18-APR-1996  
 ; ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31458-4/1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2652 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-953-492-1

Query Match 76.0%; Score 11.4; DB 2; Length 2652;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgt 14  
||||:|:|:|:|  
Db 1892 TCGAGTTGGTGT 1904

RESULT 13  
US-08-149-695-1  
Sequence 1, Application US/08149695  
Patent No. 5412085  
GENERAL INFORMATION:  
APPLICANT: Allen, Rebecca L.  
APPLICANT: Lonsdale, David M.  
TITLE OF INVENTION: A Pollen-Specific Promoter From Maize  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held and Malloy  
STREET: 500 W. Madison, 34th Floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: U.S.A.  
ZIP: 60661

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149,695  
FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/911,532  
FILING DATE: 09-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pochopien, Donald J.  
REGISTRATION NUMBER: 32167  
REFERENCE/DOCKET NUMBER: 92 P139  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)707-8889  
TELEFAX: (312)707-9155  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Zea mays  
STRAIN: Line W22

TISSUE TYPE: Pollen  
US-08-149-695-1

Query Match 76.0%; Score 11.4; DB 1; Length 2873;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cgastwtsgwgtt 15  
||||:|:|:|:|  
Db 111 CGAGTTGGAGTT 123

RESULT 14  
US-08-377-228-1  
Sequence 1, Application US/08377228  
Patent No. 5545546  
GENERAL INFORMATION:  
APPLICANT: Allen, Rebecca L.  
APPLICANT: Lonsdale, David M.  
TITLE OF INVENTION: A Pollen-Specific Promoter From Maize  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/377,228  
FILING DATE: 24-JAN-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/149,695  
FILING DATE: 09-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/911,532  
FILING DATE: 09-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 33229/290/PIHI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2873 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-377-228-1

Query Match 76.0%; Score 11.4; DB 1; Length 2873;  
Best Local Similarity 69.2%; Pred. No. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 cgastwtsgwgtt 15  
||||:|:|:|:|  
Db 111 CGAGTTGGAGTT 123

RESULT 15  
US-08-693-457-1  
Sequence 1, Application US/08693457



; Patent No. 5880330  
; GENERAL INFORMATION:  
; APPLICANT: Weigel et al., Detlef  
; TITLE OF INVENTION: SHOOT MERISTEM SPECIFIC PROMOTER  
; TITLE OF INVENTION: SEQUENCES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/693,457  
; FILING DATE: 07-AUG-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Halle, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 07251/012001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3139 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 571..1900  
US-08-693-457-1

Query Match 76.0%; Score 11.4; DB 2; Length 3139;  
Best Local Similarity 69.2%; Pred. NO. 3.6e+02;  
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
Oy 3 cgastwtsgwgtt 15  
|||:|:|:|:|  
Db 1546 CGACTTGGAGTT 1558

Search completed: February 25, 2002, 18:05:13  
Job time: 18561 sec

**THIS PAGE BLANK** (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:20:54 ; Search time 8261.74 Seconds  
(without alignments)  
19,510 Million cell updates/sec

Title: US-09-698-903B-4  
Perfect score: 15  
Sequence: 1 ntcgastwtsgwgtt 15

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_htc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_htc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
C 1	12.4	82.7 101 13 CNS04Q4L	AL302142 Tetraodon
C 2	12.4	82.7 203 11 BG199530	BG199530 RSP18921
C 3	12.4	82.7 204 10 BB258732	BB258732 BB258732
C 4	12.4	82.7 210 10 BB411504	BB411504 BB411504
C 5	12.4	82.7 216 10 AL367282	AL367282 MTBA13H05
C 6	12.4	82.7 221 10 AV427766	AV427766 AV427766
C 7	12.4	82.7 226 11 BG057582	BG057582 nah92601
C 8	12.4	82.7 247 10 AV129451	AV129451 AV129451
C 9	12.4	82.7 248 10 BB074324	BB074324 BB074324
C 10	12.4	82.7 252 10 AV331577	AV331577 AV331577
C 11	12.4	82.7 257 10 BE530468	BE530468 M77D22STM
C 12	12.4	82.7 288 13 AZ921129	AZ921129 1006024D0

13	12.4	82.7	291	10	BB191477
C 14	12.4	82.7	297	10	AV069588
C 15	12.4	82.7	298	11	BF460492
C 16	12.4	82.7	301	10	AJ284234
C 17	12.4	82.7	302	11	F14285
C 18	12.4	82.7	307	10	AV558749
C 19	12.4	82.7	307	10	BB255970
C 20	12.4	82.7	310	10	AV421222
C 21	12.4	82.7	316	10	BB250639
C 22	12.4	82.7	335	11	BF098143
C 23	12.4	82.7	342	10	AV209024
C 24	12.4	82.7	343	10	AV392544
C 25	12.4	82.7	344	10	AW693454
C 26	12.4	82.7	345	11	D40568
C 27	12.4	82.7	349	10	AV534046
C 28	12.4	82.7	355	10	AJ284258
C 29	12.4	82.7	355	10	AV441349
C 30	12.4	82.7	356	11	BF590524
C 31	12.4	82.7	359	10	AV410421
C 32	12.4	82.7	360	11	C71547
C 33	12.4	82.7	366	10	BE321781
C 34	12.4	82.7	374	11	BF778123
C 35	12.4	82.7	376	10	AA754245
C 36	12.4	82.7	376	10	AV202387
C 37	12.4	82.7	376	10	AV202461
C 38	12.4	82.7	377	10	AV422965
C 39	12.4	82.7	382	10	AV409979
C 40	12.4	82.7	382	10	AW796228
C 41	12.4	82.7	386	11	BI241776
C 42	12.4	82.7	394	10	BE192723
C 43	12.4	82.7	397	11	BF483146
C 44	12.4	82.7	405	11	BG630990
C 45	12.4	82.7	406	10	AW859801

ALIGNMENTS

RESULT	1	CNS04Q4L/C	101 bp	DNA	GSS	24-MAY-2000
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						



cap-trapper. cDNA went through one round of normalization to Rot - 20.0 and subtraction to Rot - 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATCCGCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 49 a 40 c 36 g 79 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 204;  
Best Local Similarity 71.4%; Pred. No. 3e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 4; Mismatches 0;

Oy 2 tcgastwtsgwgtt 15  
||||:|:|:|  
Db 99 TCGACTTTGGTGTT 112

RESULT 4  
BB411504/c 210 bp mRNA EST 16-JUL-2000  
LOCUS BB411504 RIKEN full-length enriched, 7 days embryo Mus musculus  
DEFINITION cDNA clone C430023B01 3', mRNA sequence.

ACCESSION BB411504  
VERSION BB411504.1 GI:9232859  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 210)  
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers  
1. .210

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="C430023B01"  
/clone\_lib="RIKEN full-length enriched, 7 days embryo"  
/dev\_stage="7 days embryo"  
/lab\_host="DH10B"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATCCGCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 63 a 54 c 41 g 52 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 210;  
Best Local Similarity 71.4%; Pred. No. 3e+03; Indels 0; Gaps 0;  
Matches 10; Conservative 4; Mismatches 0;

Oy 2 tcgastwtsgwgtt 15  
||||:|:|:|  
Db 18 TCGAGTATCGTGTT 5

RESULT 5  
AL367282/c 216 bp mRNA EST 03-AUG-2000  
LOCUS AL367282 MTBAl3H05F1 MCBAl3H05F1 cDNA clone MCBAl3H05 T3, mRNA  
DEFINITION sequence.

ACCESSION AL367282.1 GI:9667035  
VERSION AL367282  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE 1 (bases 1 to 216)  
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.  
Medicago truncatula ESTs from nitrogen-starved roots  
Unpublished (2000)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91005 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

CONTACT : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).  
Location/Qualifiers  
1. .216

FEATURES source  
1. .216  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/db\_xref="taxon:3880"  
/clone="MTBAl3H05"  
/clone\_lib="MtBA"  
/tissue\_type="root tips"  
/note="Vector: pBluescript pSK; Site\_1: EcoRI; Site\_2: "

XhoI; Plants were grown in an aeroponic chamber for 14 days on nitrogen-rich medium followed by 3 days on N-free medium. RNA was extracted from root tips (1-3 cm). cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapXR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequençage (Genoscope, Evry, France)."

BASE COUNT 69 a 56 c 34 g 56 t 1 others  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 216;  
Best Local Similarity 71.4%; Pred. No. 3e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
||||:|:|:|:|

Db 74 TCGAGTTTCGAGTT 61

RESULT 6  
AV427766/c

LOCUS AV427766 221 bp mRNA EST 23-MAY-2000  
DEFINITION japonicus cDNA clone MM086e02\_r 5', mRNA sequence.  
ACCESSION AV427766  
VERSION AV427766.1 GI:7788042  
KEYWORDS EST.  
SOURCE Lotus japonicus.  
ORGANISM Lotus japonicus.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.

REFERENCE 1 (bases 1 to 221)  
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus  
DNA Res. 7 (2), 127-130 (2000)

JOURNAL 20277479  
MEDLINE  
COMMENT Contact: Yasukazu Nakamura  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1..221  
/organism="Lotus japonicus"  
/db\_xref="taxon:34305"  
/clone\_lib="Lotus japonicus young plants (two-week old)"  
/dev\_stage="young plants (two-week old)"  
/notes="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI; isolate=MiyakoJima MG-20"

BASE COUNT 73 a 60 c 25 g 63 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 221;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
||||:|:|:|:|

Db 49 TCGAGTTTCGAGTT 36

RESULT 7  
BG057582

LOCUS BG057582 226 bp mRNA EST 25-JAN-2001  
DEFINITION nah92d01.x1 NCI\_CGAP\_HN17 Homo sapiens cDNA clone IMAGE:4258200 3', mRNA sequence.  
ACCESSION BG057582  
VERSION BG057582.1 GI:12523216  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 226)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
unknown library type  
Seq primer: -400P from Gibco.  
FEATURES  
source  
1..226  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:4258200"  
/clone\_lib="NCI\_CGAP\_HN17"  
/tissue\_type="normal epithelium"  
/lab\_host="DH10B"  
/note="Organ: nasopharynx; Vector: pAMP10; mRNA made from normal nasopharyngeal epithelium, cDNA made by oligo-dT priming. Non-directionally cloned into UDG sites.  
Size-selected on agarose gel, average insert size 500 bp.  
Primary library. cDNA Library Preparation: David B. Krizman, Ph.D. REFERENCE: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 63 a 59 c 43 g 61 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 11; Length 226;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
||||:|:|:|:|

Db 2 TCGACTTTGGTGT 15

RESULT 8  
AV129451/c

LOCUS AV129451 247 bp mRNA EST 01-JUL-1999  
DEFINITION AV129451 Mus musculus C57BL/6J 11-day embryo Mus musculus cDNA clone 2700068H03, mRNA sequence.  
ACCESSION AV129451  
VERSION AV129451.1 GI:5315686  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 247)  
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Alizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Suganara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs  
JOURNAL Unpublished (1999)  
COMMENT Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp  
Thermotransformation and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

#### FEATURES

source  
Location/Qualifiers  
1. .247  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2700068H03"  
/clone\_lib="Mus musculus C57BL/6J 11-day embryo"  
/sex="mixed"  
/dev\_stage="11-day embryo"  
77 a 41 c 34 g 95 t  
BASE COUNT  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 247;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15  
||||:|:|:|:|:|

Db 121 TCGAGTTTGGTGTT 108

#### RESULT 9

BB074324 248 bp mRNA EST 27-JUN-2000  
LOCUS BB074324 RIKEN full-length enriched, adult male diencephalon Mus  
DEFINITION musculus cDNA clone 9330013E13 3' similar to S78234 nuc2 homolog,  
mRNA sequence.

#### ACCESSION

BB074324.1 GI:8584322

#### VERSION

EST.

#### KEYWORDS

EST.

#### SOURCE

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 248)

REFERENCE

AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci

,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,

Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,

Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,

Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,

Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,

Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata

,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,

Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,

Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino

,M., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Konno,H., et al.)

Unpublished (2000)

CONTACT: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@rtc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagasaka,S., Sasaki

,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermotransformation and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,  
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki  
,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

#### FEATURES

source  
Location/Qualifiers  
1. .248  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="9330013E13"  
/clone\_lib="RIKEN full-length enriched, adult male  
diencephalon"  
/sex="male"  
/tissue\_type="diencephalon"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 185.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATTCGAGTTAATTAATCCGCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pBluescript KS(+) after bulk excision  
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:  
BamHI"

BASE COUNT 68 a 31 c 50 g 99 t

#### ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 248;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 2 tcgastwtsgwgtt 15

||||:|:|:|:|:|

Db 197 TCGAGTATGGTGTT 210

#### RESULT 10

AV331577

LOCUS AV331577

DEFINITION AV331577 RIKEN full-length enriched, adult male medulla oblongata

Mus musculus cDNA clone 6330524M18 3', mRNA sequence.

AV331577

ACCESSION AV331577.1 GI:6371629

VERSION EST.

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 252)

REFERENCE

AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,

Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,

Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai

,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,

Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,

Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata

,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,

Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya

Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Taten,M., Tomimaga,N., Tsunoda,Y., Yashiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al. 1999)  
Unpublished (1999)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh,M., Kitsumai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers  
1. .252

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="630524M18"  
/clone\_lib="RIKEN full-length enriched, adult male medulla oblongata"  
/sex="male"  
/tissue\_type="medulla oblongata"  
/lab\_host="adult"  
/lab\_host="DH108"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 67 a 53 c 53 g 79 t  
ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 252;  
Best Local Similarity 71.4%; Pred. No. 3.2e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|||||

Db 129 TCGACTTGGTGT 142

RESULT 11  
BE530468/c

LOCUS

DEFINITION

BE530468 257 bp mRNA EST 19-MAR-2001  
M77D22STM Arabidopsis developing seed Arabidopsis thaliana cDNA clone 600039982R1 5', mRNA sequence.

ACCESSION

BE530468

VERSION

BE530468.1 GI:9788458

KEYWORDS

EST.

SOURCE

ORGANISM

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 257)

White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de

Illarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.

A new set of Arabidopsis expressed sequence tags from developing

seeds. The metabolic pathway from carbohydrates to seed oil

Plant Physiol. 124 (4), 1582-1594 (2000)

20567808

COMMENT

Contact: Benning, C

Dept. of Biochemistry & Molecular Biology

Michigan State University

224 Biochemistry, Michigan State University, East Lansing, MI 48824

, USA

Tel: 517 355 1609

Fax: 517 353 9334

Email: benning@msu.edu

Clones were originally prepared at Michigan State University.

Arabidopsis Biological Resource Center, The Ohio State University,

309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210

USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

source

Location/Qualifiers

1. .257

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="600039982R1"

/clone\_lib="Arabidopsis developing seed"

/tissue\_type="seed"

/dev\_stage="5-13 days after flowering"

/lab\_host="E.coli"

/note="Organ: Developing seed; Vector: pBluescript SK-;

Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 85 a 71 c 35 g 66 t

ORIGIN

Query Match 82.7%; Score 12.4; DB 10; Length 257;

Best Local Similarity 71.4%; Pred. No. 3.2e+03;

Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15

||||:|||||

Db 58 TCGAGTTTCGTGT 45

RESULT 12

AZ921129/c

LOCUS

AZ921129

DEFINITION

1006024D02.y1 1006 - RescueMu Grid G Zea mays genomic, DNA

sequence.

ACCESSION

AZ921129

VERSION

AZ921129.1 GI:13392455

KEYWORDS

GSS.

SOURCE

Zea mays.

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 288)

AUTHORS

Walbot,V.





# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AV069588  
AV069588.1 GI:5189416  
EST.  
house mouse.  
Mus musculus

## REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 297)  
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,  
A., Hayatsu, N., Hori, F., Ishikawa, T., Fukunishi, Y., Funayama, T., Hara  
Kikuchi, N., Kojima, Y., Matsuyama, T., Itoh, M., Izawa, M., Kawai, J.,  
Sato, K., Shibata, Y., Shigemoto, Y., Nitsuma, H., Oda, H., Owa, C.,  
Y., Suzuki, H., Suzuki, H., Tatenno, M., Shiraki, T., Sogabe, Y., Sugahara  
Watanabe, S., Yagame, M., Yamamura, T., Tomaru, Y., Tominaga, N.,  
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

## TITLE JOURNAL COMMENT

Unpublished (1999)  
Contact: Chie Owa  
Genome Science Laboratory  
RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098  
Email: genome-resetc.riken.go.jp

Thermotabilization and thermoactivation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length cDNA  
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3453-3460 (1998))  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES source

1..297  
/location="Mus musculus"  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="2010310L11"  
/clone\_lib="Mus musculus small intestine C57BL/6J adult"  
/sex="male"  
/tissue\_type="small intestine"  
/dev\_stage="adult"  
95 a 89 c 68 g 45 t

Query Match 82.7%; Score 12.4; DB 10; Length 297;  
Best Local Similarity 71.4%; Pred. No. 3.3e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
|||||:|:|:|:|  
Db 74 TCGAGTTTGGTGT 61

## RESULT 15 BF460492/c LOCUS

DEFINITION BP460492 298 bp mRNA EST 04-DEC-2000  
UI-M-CGOp-bmd-c-02-0-UI.s1 NIH\_BMAP\_Ret4\_S2 Mus musculus cDNA clone  
BF460492  
BF460492 GI:11529649

## ACCESSION VERSION KEYWORDS SOURCE ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 298)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477

# COMMENT

Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: m5t5@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
retina tissue cDNA Library Preparation: M.B. Soares Lab Clone  
distribution: Researchers may obtain BMAP cDNA clones from RESEARCH  
GENETICS. It should be noted that Bento Soares is generating a  
small number of additional specialized non-redundant arrays of BMAP  
cDNAs whose availability will be considered under appropriate and  
limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES source

Location/Qualifiers  
1..298  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CGOp-bmd-c-02-0-UI"  
/clone\_lib="NIH\_BMAP\_Ret4\_S2"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker: Site.1: Not I; Site.2: Eco RI; The  
NIH\_BMAP\_Ret4\_S2 library is a subtracted library,  
ultimately derived from mouse retina tissue libraries at  
various stages of development. For a detailed description  
of the library from which this clone was derived, please  
visit our web site at [brainest.eng.uiowa.edu](http://brainest.eng.uiowa.edu).  
TAG\_LIB=NIH\_BMAP\_Ret4\_S2  
TAG\_TISSUE=adult-retina  
TAG\_SEQ=GTACGCGGCAC"  
73 a 74 c 77 g 74 t

Query Match 82.7%; Score 12.4; DB 11; Length 298;  
Best Local Similarity 71.4%; Pred. No. 3.3e+03;  
Matches 10; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgastwtsgwgtt 15  
|||||:|:|:|:|  
Db 121 TCGAGTTTGGAGTT 108

Search completed: February 25, 2002, 17:20:57  
Job time: 16150 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:08 ; Search time 2331.3 Seconds  
(without alignments)  
176.910 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25

Sequence: 1 ggatcccccgatgagtaagctagc 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	6	AX127752 Sequence
2	25	100.0	25	6	AX172446 Sequence
3	25	100.0	249	12	ARGMTUB
4	25	100.0	1037	6	AX10942
5	25	100.0	1085	6	AX10939
6	25	100.0	1160	6	AX10943
c 7	25	100.0	2476	12	TBI251013
c 8	25	100.0	3236	12	TBI251014
c 9	25	100.0	4832	6	AX172441
c 10	25	100.0	4946	6	AX172441
c 11	25	100.0	4946	6	AX172441
c 12	25	100.0	4946	6	AX172441
c 13	25	100.0	4946	6	AX172441
c 14	25	100.0	5349	6	AX172441
c 15	25	100.0	5560	6	AX172441
c 16	25	100.0	5560	6	AX172441
c 17	25	100.0	5865	6	AX172441
c 18	25	100.0	5865	6	AX172441
c 19	25	100.0	6539	6	AX172441
c 20	25	100.0	6548	6	AX172441
c 21	25	100.0	6548	6	AX172441
c 22	25	100.0	6548	6	AX172441
c 23	25	100.0	6548	6	AX172441
c 24	25	100.0	7599	6	AX172441
c 25	25	100.0	12095	12	BINHYGDA
c 26	24	96.0	1166	6	AX10941
c 27	24	96.0	3200	6	AX10941
c 28	24	96.0	3201	6	AX10941
c 29	24	96.0	7566	6	AX10941
c 30	24	96.0	7566	6	AX10941
c 31	24	96.0	7639	6	AX10941
c 32	24	96.0	7639	6	AX10941
c 33	24	96.0	7811	6	AX10941
c 34	19.2	76.8	134832	2	AX10941
c 35	18.6	75.2	152883	8	AX10941
c 36	18.6	74.4	175826	2	AX10941
c 37	17.8	71.2	25514	2	AX10941
c 38	17.8	71.2	175118	3	AX10941
c 39	17.8	71.2	188272	3	AX10941
c 40	17.8	71.2	298616	3	AX10941
c 41	17.6	70.4	83078	8	AX10941
c 42	17.6	70.4	93489	8	AX10941
c 43	17.6	70.4	149172	9	AX10941
c 44	17.6	70.4	149930	9	AX10941
c 45	17.6	70.4	160570	2	AX10941

ALIGNMENTS

RESULT 1

AX127752

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

AX127752

Sequence 5 from Patent WO0131042.

AX127752

AX127752.1 GI:14134399

synthetic construct.

synthetic construct

artificial sequence.

1 (bases 1 to 25)

Weston, B. and de Beuckeleer, M.

Male-sterile brassica plants and methods for producing same

Patent: WO 0131042-A 5 03-MAY-2001;

Aventis CropScience N.V. (BE)

Location/Qualifiers

1..25

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer M8251"

15-MAY-2001

PAT

```

BASE COUNT      6 a      8 c      7 g      4 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
|||||
Db 1 GGATCCCCCGATGAGCTAAGCTAGC 25

RESULT 2
AXI72446
LOCUS      AXI72446      25 bp      DNA
DEFINITION Sequence 7 from Patent WO0141558.
ACCESSION AXI72446
VERSION    AXI72446.1 GI:14597558
KEYWORDS   synthetic construct.
SOURCE     synthetic construct
ORGANISM   artificial sequence.
REFERENCE  1 (bases 1 to 25)
AUTHORS    de Both,G. and de Beuckeleer,M.
TITLE      Hybrid winter oilseed rape and methods for producing same
JOURNAL    Patent: WO 0141558-A 7 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES   Location/Qualifiers
            source
            1..25
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="primer 231"

BASE COUNT      6 a      8 c      7 g      4 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
|||||
Db 1 GGATCCCCCGATGAGCTAAGCTAGC 25

RESULT 3
ARGMTUB
LOCUS      ARGMTUB      249 bp      DNA
DEFINITION Soybean beta-1-tubulin gene fused to Ti plasmid unit 7 30TR.
ACCESSION X05579
VERSION    X05579.1 GI:58087
KEYWORDS   beta-tubulin; fusion gene; plasmid.
SOURCE     synthetic construct.
ORGANISM   synthetic construct
REFERENCE  1 (bases 1 to 249)
AUTHORS    Guiltinan,M.J., Velten,J., Bustos,M.M., Cyr,R.J., Schell,J. and
Fosket,D.E.
TITLE      The expression of a chimeric soybean beta-tubulin gene in tobacco
JOURNAL    Mol. Gen. Genet. 207, 328-334 (1987)
FEATURES   Location/Qualifiers
            source
            1..249
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="fusion product (17AA); Protein sequence is in
            conflict with the conceptual translation"
            /codon_start=1
            /transl_table=11
            /protein_id="CAA29084.1"
            /db_xref="GI:4376141"
            /translation="AMAAWASSNWSTDPMS"

CDS

```

---

```

misc_feature      1..13
                  /note="beta-1-tubulin sequence"
misc_feature      14..28
                  /note="pUC 13 polylinker"
misc_feature      29..35
                  /note="Sal I linker"
misc_feature      37
                  /note="theroretical fusion junction (24) with gene 7 of Ti
                  plasmid"
misc_feature      173..178
                  /note="put.polyA signal"
polyA_site        198
                  /note="polyA site"
misc_feature      220..225
                  /note="put.polyA signal"

BASE COUNT      76 a      47 c      38 g      88 t
ORIGIN

Query Match      100.0%; Score 25; DB 12; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
|||||
Db 36 GGATCCCCCGATGAGCTAAGCTAGC 60

RESULT 4
AL0942
LOCUS      AL0942      1037 bp      DNA
DEFINITION Nucleotide sequence 4 from patent number DE3920034.
ACCESSION AL0942
VERSION    AL0942.1 GI:492369
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unidentified
REFERENCE  1 (bases 1 to 1037)
AUTHORS    .
JOURNAL    Patent: DE 3920034-A 4 31-MAY-1990;
FEATURES   Location/Qualifiers
            source
            1..1037
            /organism="unidentified"
            /db_xref="taxon:32644"

BASE COUNT      338 a      174 c      166 g      359 t
ORIGIN

Query Match      100.0%; Score 25; DB 6; Length 1037;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
|||||
Db 767 GGATCCCCCGATGAGCTAAGCTAGC 791

RESULT 5
AL0939
LOCUS      AL0939      1085 bp      DNA
DEFINITION Nucleotide sequence 1 from patent number DE3920034.
ACCESSION AL0939
VERSION    AL0939.1 GI:492367
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unidentified
REFERENCE  1 (bases 1 to 1085)
AUTHORS    .
JOURNAL    Patent: DE 3920034-A 1 31-MAY-1990;
FEATURES   Location/Qualifiers
            source
            1..1085

```

```
/organism="unidentified"
/db_xref="taxon:32644"
369 a 218 c 155 g 343 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1085;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagtacg 25
|||||
Db 815 GGATCCCCCGATGAGCTAAGCTAGC 839

RESULT 6
LOCUS A10943 1160 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 5 from patent number DE3920034.
ACCESSION A10943
VERSION A10943.1 GI:492370
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1160)
AUTHORS
JOURNAL Patent: DE 3920034-A 5 31-MAY-1990;
FEATURES
source Location/Qualifiers
1..1160
/organism="unidentified"
/db_xref="taxon:32644"
367 a 194 c 188 g 411 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagtacg 25
|||||
Db 890 GGATCCCCCGATGAGCTAAGCTAGC 914

RESULT 7
TBI251013/c
LOCUS TBI251013 2476 bp DNA SYN 26-JUN-2001
DEFINITION Transformation binary vector pBAR-A, T-DNA region.
ACCESSION AJ251013
VERSION AJ251013.1 GI:6453666
KEYWORDS t-dna; transformation binary vector.
SOURCE Transformation binary vector pBAR-A.
ORGANISM artificial sequence; vectors.
REFERENCE 1 (bases 1 to 2476)
AUTHORS Cardon,G.H. and Huijser,P.
TITLE T-DNA sequence of the plant transformation binary vector pBAR-A
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2476)
AUTHORS Cardon,G.H.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-1999) Cardon G.H., Molecular Plant Genetics,
Max-Planck- Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg
10, 50829 Cologne, GERMANY
COMMENT Related sequences: U09365, J01825, J01636, X02513, V00090, X05822,
X01077, J01826.
FEATURES
source Location/Qualifiers
1..2476
/organism="Transformation binary vector pBAR-A"
/db_xref="taxon:108140"
/note="can be cultured in Escherichia coli or
Agrobacterium tumefaciens-T-DNA region of a binary vector

derived from plasmid pgPTV-BAR"
1..148
/note="similar to T-DNA left border"
complement(149..279)
/gene="laci"
CDS complement(<149..>279)
/gene="laci"
/codon_start=1
/transl_table=11
/protein_id="CAC42834.1"
/db_xref="GI:14572627"
/translation="NHHQTGFSPAGANQRGPIAATLSCPGGEGQSAAVARLTGEEKNHHP"
280..576
/note="M13 ori"
/direction=RIGHT
complement(<577..>732)
/gene="lac2"
complement(577..732)
/gene="lac2"
733..955
/gene="transcript 7"
733..955
/gene="transcript 7"
complement(<956..>1544)
/gene="bar"
complement(956..1544)
/gene="bar"
1545..1837
/gene="NOS"
1545..2314
/gene="NOS"
1838..2157
/gene="NOS"
/note="similar to transposable element Tn5"
2158..2191
/gene="NOS"
/note="synthetic multiple cloning site"
2192..2314
/gene="NOS"
2315..2476
/note="similar to T-DNA right border"
BASE COUNT 563 a 614 c 699 g 600 t
ORIGIN

Query Match 100.0%; Score 25; DB 12; Length 2476;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagtacg 25
|||||
Db 961 GGATCCCCCGATGAGCTAAGCTAGC 937

RESULT 8
TBI251014/c
LOCUS TBI251014 3236 bp DNA SYN 26-JUN-2001
DEFINITION Transformation binary vector pBAR-35S, T-DNA region.
ACCESSION AJ251014
VERSION AJ251014.1 GI:6453667
KEYWORDS t-dna; transformation binary vector.
SOURCE Transformation binary vector pBAR-35S.
ORGANISM Transformation binary vector pBAR-35S.
REFERENCE 1 (bases 1 to 3236)
AUTHORS Cardon,G.H. and Huijser,P.
TITLE T-DNA sequence of the plant transformation binary vector pBAR-A
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3236)
AUTHORS Cardon,G.H.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-1999) Cardon G.H., Molecular Plant Genetics,
```

Max-Planck- Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg  
10, 50829 Cologne, GERMANY  
Related sequences: U09365, J01825, J01636, X02513, V00090, X05822  
X01077, J01826.

FEATURES  
    source  
        1..3236  
            /organism="Transformation binary vector pBAR-35S"  
            /db\_xref="taxon:108141"  
            /note="Can be cultured in Escherichia coli or  
            Agrobacterium tumefaciens-T-DNA region of a binary vector  
            derived from plasmid pGPTV-BAR"  
    misc\_feature  
        1..148  
            /note="similar to T-DNA left border"  
    gene  
        complement(149..279)  
        /gene="lacI"  
    CDS  
        complement(<149..>279)  
        /gene="lacI"  
        /codon\_start=1  
        /transl\_table=11  
        /protein\_id="CAC42835.1"  
        /db\_xref="GI:14572628"  
        /translation="NHQTGFSPAGANORGLAATLSPPGGEGQSAVARLTGKKNHP  
        \*  
        280..576  
            /note="M13 ori"  
            complement(<577..>732)  
            /gene="lacZ"  
            complement(577..732)  
            /gene="lacZ"  
            733..955  
            /gene="transcript 7"  
            733..955  
            /gene="transcript 7"  
            /note="putative"  
            complement(<956..>1544)  
            /gene="bar"  
            complement(956..1544)  
            /gene="bar"  
            1545..1837  
            /gene="NOS"  
            1545..1837  
            /gene="NOS"  
            1838..2157  
            /note="similar to transposable element Tn5"  
            2167..2707  
            /gene="camV 35S"  
            2167..2256  
            /gene="camV 35S"  
            2708..2730  
            /gene="camV 35S"  
            /note="Synthetic multiple cloning site"  
            2731..2956  
            /gene="camV 35S"  
            2957..3074  
            /gene="NOS gene promoter"  
            2957..3074  
            /gene="NOS gene promoter"  
            3075..3236  
            /note="similar to T-DNA right border"  
BASE COUNT   804 a   786 c   860 g   786 t  
ORIGIN

Query Match           100.0%; Score 25; DB 12; Length 3236;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches   25; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

QY   1   ggatccccgatgactaagctagc 25  
      |||||  
Db   961   GGATCCCCCGATGAGCTAAGCTAGC 937

RESULT   9

AX172441/c  
LOCUS                   AX172441           4832 bp   DNA  
DEFINITION            Sequence 2 from Patent WO0141558.  
ACCESSION            AX172441  
VERSION               AX172441.1   GI:14597553  
KEYWORDS  
SOURCE                synthetic construct.  
ORGANISM             artificial construct  
                      artificial sequence.  
REFERENCE            1 (bases 1 to 4832)  
                      de Both.G. and de Beuckeleer, M.  
                      Hybrid winter oilseed rape and methods for producing same  
                      Patent: WO 0141558-A 2 14-JUN-2001;  
                      Aventis CropScience N.V. (BE)  
TITLE                 Location/Qualifiers  
JOURNAL  
FEATURES  
    source  
        1..4832  
            /organism="synthetic construct"  
            /db\_xref="taxon:32630"  
            /note="T-DNA of plasmid pTHW118"  
    misc\_feature  
        1883..4065  
            /note="HpaI restriction fragment"  
BASE COUNT   1528 a   883 c   932 g   1488 t  
ORIGIN

Query Match           100.0%; Score 25; DB 6; Length 4832;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches   25; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

QY   1   ggatccccgatgactaagctagc 25  
      |||||  
Db   317   GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT   10

A60108/c  
LOCUS                   A60108           4946 bp   DNA  
DEFINITION            Sequence 1 from Patent WO9706267.  
ACCESSION            A60108  
VERSION               A60108.1   GI:3715124  
KEYWORDS  
SOURCE                Transformation vector pTHW107.  
ORGANISM             Transformation vector pTHW107  
                      artificial sequence; vectors.  
REFERENCE            1 (bases 1 to 4946)  
AUTHORS             De B.M.  
TITLE                 Genetic transformation using a PARP inhibitor  
JOURNAL             Patent: WO 9706267-A 1 20-FEB-1997;  
                      PLANT GENETIC SYSTEMS NV (BE)  
FEATURES  
    source  
        1..4946  
            /organism="Transformation vector pTHW107"  
            /db\_xref="taxon:126810"  
BASE COUNT   1569 a   891 c   963 g   1523 t  
ORIGIN

Query Match           100.0%; Score 25; DB 6; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches   25; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

QY   1   ggatccccgatgactaagctagc 25  
      |||||  
Db   317   GGATCCCCCGATGAGCTAAGCTAGC 293

RESULT   11

A76915/c  
LOCUS                   A76915           4946 bp   DNA  
DEFINITION            Sequence 1 from Patent EP0757102.  
ACCESSION            A76915  
VERSION               A76915.1   GI:6088712  
KEYWORDS

SOURCE Transformation vector pTHW107.  
ORGANISM Transformation vector pTHW107  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De,B.M.  
TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR  
JOURNAL Patent: EP 0757102-A 1 05-FEB-1997;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="Transformation vector pTHW107"  
/db\_xref="taxon:126810"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25  
|||||  
Db 317 GGATCCCCCGATGACTAGCTAGC 293

RESULT 12  
AR098307/c  
LOCUS AR098307 4946 bp DNA 14-FEB-2001  
DEFINITION Sequence 1 from patent US 6074876.  
ACCESSION AR098307  
VERSION AR098307.1 GI:12807564  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De Block,M.  
TITLE Genetic transformation using a PARP inhibitor  
JOURNAL Patent: US 6074876-A 1 13-JUN-2000;  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="unknown"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25  
|||||  
Db 317 GGATCCCCCGATGACTAGCTAGC 293

RESULT 13  
AX172440/c  
LOCUS AX172440 4946 bp DNA 03-JUL-2001  
DEFINITION Sequence 1 from Patent WO0141558.  
ACCESSION AX172440  
VERSION AX172440.1 GI:14597552  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS de Both,G. and de Beuckeleer,M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 1 14-JUN-2001;  
Aventis CropScience N.V. (BE)  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="synthetic construct"

/db\_xref="taxon:32630"  
/note="T-DNA of plasmid pTHW107"  
964. .4906  
/note="Hind III fragment"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25  
|||||  
Db 317 GGATCCCCCGATGACTAGCTAGC 293

RESULT 14  
A71437/c  
LOCUS A71437 5349 bp DNA 07-MAY-1999  
DEFINITION Sequence 7 from Patent WO9810081.  
ACCESSION A71437  
VERSION A71437.1 GI:4775050  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 5349)  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 7 12-MAR-1998;  
MICHIELS FRANK (BE)  
FEATURES Location/Qualifiers  
source  
1. 5349  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 1339 a 1233 c 1290 g 1487 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5349;  
Best Local Similarity 100.0%; Pred. No. 0.035;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25  
|||||  
Db 317 GGATCCCCCGATGACTAGCTAGC 293

RESULT 15  
A60112/c  
LOCUS A60112 5560 bp DNA 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9706267.  
ACCESSION A60112  
VERSION A60112.1 GI:3715128  
KEYWORDS  
SOURCE Transformation vector pTHW142.  
ORGANISM Transformation vector pTHW142.  
REFERENCE 1 (bases 1 to 5560)  
AUTHORS De,B.M.  
TITLE Genetic transformation using a PARP inhibitor  
JOURNAL Patent: WO 9706267-A 5 20-FEB-1997;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES Location/Qualifiers  
source  
1. .5560  
/organism="Transformation vector pTHW142"  
/db\_xref="taxon:126817"  
BASE COUNT 1533 a 1199 c 1244 g 1576 t  
ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 5560;





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:22 ; Search time 716.55 Seconds  
(without alignments)  
29.912 Million cell updates/sec

Title: US-09-698-903B-5  
Perfect score: 25  
Sequence: 1 ggtatccccgatgagtaagctagc 25  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	AAH25426	PCR primer for fla
2	25	100.0	25	AAH25426	PCR primer MDB251
3	25	100.0	1037	AAH25426	USP-Promoter-casse
4	25	100.0	1085	AAH25426	Legumin-signalpept
5	25	100.0	1160	AAH25426	USP-signalpeptide
6	25	100.0	1303	AAH25426	Plasmid pTS88 (Eco
7	25	100.0	4832	AAH25426	Nucleotide sequenc
8	25	100.0	4946	AAH25426	T-DNA of plasmid p
9	25	100.0	4946	AAH25426	Nucleotide sequenc
10	25	100.0	5228	AAH25426	Plasmid pTS172delt
11	25	100.0	5349	AAH25426	T-DNA of pTTS24.

12	25	100.0	5864	17	AAT39339	Plasmid pTColl13 T-
13	25	100.0	5864	17	AAT39339	Chimeric T-DNA of
14	25	100.0	5865	22	AAH25426	Chimeric T-DNA of
15	25	100.0	5865	22	AAH25426	E. coli plasmid pT
16	25	100.0	6539	21	AAZ91097	Plasmid pTS174 use
17	25	100.0	6548	17	AAT39336	Plasmid pTS172. C
18	25	100.0	6548	18	AAT61394	E. coli plasmid pT
19	25	100.0	6548	21	AAZ91096	Plasmid pTS346. U
20	25	100.0	7492	22	AAF86441	Nucleotide sequenc
21	25	100.0	7599	22	AAZ91096	USP-Promoter-casse
22	24	96.0	1166	11	AAQ04704	PP5029 Bt ICP codi
23	24	96.0	3201	12	AAQ14529	PVE36 Bt ICP codi
24	24	96.0	3201	12	AAQ15144	Plasmid pPS0212 co
25	24	96.0	7566	14	AAQ42160	Plasmid pPS0212 co
26	24	96.0	7639	14	AAQ42159	Agrobacterium tune
27	19	76.0	34	22	AAH43929	PCR primer used to
28	19	76.0	34	22	AAH25165	Human polynucleoti
29	17.2	68.8	2562	22	AAI59154	Human polynucleoti
30	17.2	68.8	6380	22	AAI57822	Human polynucleoti
31	17.2	68.8	7104	22	AAI59608	Human polynucleoti
32	17	68.0	29	14	AAQ42148	Bt884 oligonucleot
33	17	68.0	3153	21	AAZ29122	Plasmid DV131 comp
34	17	68.0	3336	21	AAZ29121	Plasmid DV130 comp
35	17	68.0	3694	21	AAZ29124	Plasmid DV133 used
36	17	68.0	3877	21	AAZ29123	Sequence of opine
37	17	68.0	24593	6	AAH50226	Complete nucleotid
38	17	68.0	24596	6	AAH50182	S. pneumoniae deri
39	16.6	66.4	3488	19	AAZ96307	Streptococcus pneu
40	16.6	66.4	6204	19	AAV44592	Rhizobium species
41	16.6	66.4	534720	19	AAV30458	Rhizobium species
42	16.6	66.4	536165	19	AAV30459	Arabidopsis thalia
43	16.2	64.8	457	21	AAZ53867	Arabidopsis thalia
44	16.2	64.8	521	21	AAZ54447	Arabidopsis thalia
45	16.2	64.8	526	21	AAZ52318	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAH25426	AAH25426 standard; DNA; 25 BP.
ID	AAH25426 standard; DNA; 25 BP.
XX	AC
XX	AAH25426;
XX	AC
DT	22-AUG-2001 (first entry)
XX	PCR primer for flanking regions in transgenic plant MS-BN1.
DE	Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX	fertility restorer gene; barnase gene; PCR primer; ss.
KW	Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX	fertility restorer gene; barnase gene; PCR primer; ss.
OS	Synthetic.
XX	WO200141558-A1.
PN	14-JUN-2001.
PD	06-DEC-2000; 2000WO-EP12872.
PF	08-DEC-1999; 99US-0457037.
XX	(AVET ) AVENTIS CROPS SCIENCE NV.
PA	De Both G, De Beuckeleer M;
XX	WPI; 2001-381419/40.
XX	Transgenic winter oilseed rape plants suited for producing hybrid seed
PT	with improved qualities, comprises a male-sterility gene and fertility
PT	restorer gene, integrated into the genome
XX	Example 4; Page 42; 98pp; English.
PS	

XX The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC PCR primers AAF25425-26 were used to amplify the flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barnase  
 CC transgene.

XX  
 SQ Sequence 25 BP; 6 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaaagctagc 25  
 |||||  
 Db 1 ggatccccgatgagctaaagctagc 25

RESULT 2  
 ID AAD06994 standard; DNA; 25 BP.  
 AC AAD06994;  
 DT 06-AUG-2001 (first entry)  
 DE PCR primer MDB251 to generate the flanking region of elite event MS-B2.  
 KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; thermal asymmetric interlaced;  
 KW TAIL; ss.  
 OS Agrobacterium sp.  
 PI WO200131042-A2.  
 PN 03-MAY-2001.  
 PD 26-OCT-2000; 2000WO-EPI0680.  
 PF 29-OCT-1999; 99US-0430497.  
 PR (AVET ) AVENTIS CROPS SCIENCE NV.  
 PA Weston B, De Beuckeleer M;  
 PI WPI; 2001-300517/31.  
 DR Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 XX harboring specific transformation events, particularly by presence of  
 XX male-sterility gene, at specific location in its genome -  
 XX Example 3; Page 28; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is primary thermal interlaced (TAIL)-PCR primer  
 CC MDB251 used to right (5') and left (3') border flanking region of elite  
 CC event MS-B2. This primer corresponds to position 293-317 of plasmid  
 CC pTCO113.

XX  
 SQ Sequence 25 BP; 6 A; 8 C; 7 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 22; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaaagctagc 25  
 |||||  
 Db 1 ggatccccgatgagctaaagctagc 25

RESULT 3  
 ID AAQ04705 standard; DNA; 1037 BP.  
 AC AAQ04705;  
 DT 12-OCT-1990 (first entry)  
 DE USP-Promoter-cassette USP-Pr.T7.1.  
 KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.  
 PN DE3920034-A.  
 XX 31-MAY-1990.  
 XX 20-JUN-1989; 89DE-3920034.  
 XX 19-SEP-1988; 88DD-0319887.  
 XX (PFLA-) VE KOMB PFLANZENZUC.  
 XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;  
 XX WPI; 1990-172459/23.  
 XX Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.  
 XX Disclosure; ; pp; German.  
 XX The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
 CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the TI-vector pGA471. *Agrobacterium tumefaciens* is  
 CC transfected.  
 CC See also AAQ04703-Q04706.  
 XX  
 SQ Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 100.0%; Score 25; DB 11; Length 1037;  
 Best Local Similarity 100.0%; Pred. No. 0.007;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaaagctagc 25  
 |||||  
 Db 767 ggatccccgatgagctaaagctagc 791

RESULT 4  
 ID AAQ04703 standard; DNA; 1085 BP.  
 AC AAQ04703;  
 XX 12-OCT-1990 (first entry)  
 DE Legumin-signalpeptide cassette Le-Sig.T7.  
 XX

```
KW Foreign DNA incorporation; recombinant DNA techniques;
KW higher plant genome; legumin; B4 gene; signalpeptide; Le-Sig.T7.; ss.
FH Key Location/Qualifiers
FT CDS 747..814
FT /*tag= a
FT /product=Legumin-signalpeptide
XX
XX
XX DE3920034-A.
XX
XX 31-MAY-1990.
XX
XX 20-JUN-1989; 89DE-3920034.
XX
XX 19-SEP-1988; 88DD-0319887.
XX
XX (PFLA-) VE KOMB PFLANZENZUC.
XX
XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX WPI; 1990-172459/23.
XX P-PSDB; AAR05198.
XX
XX Incorporation of DNA into higher plant genome - by specified
XX recombinant DNA techniques.
XX
XX Disclosure; ; pp; German.
XX
XX The Legumin gene B4 is used. The unique BglII-Ort (815-820) site is for
XX ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX (1080-1085) for cloning the cassette in the Ti-vector pGA471.
XX The cassette is cloned into the binary Ti-vectors pGA471 and
XX Agrobacterium tumefaciens is transfected.
XX See also AAQ04703-Q04706.
XX
XX Sequence 1085 BP; 368 A; 219 C; 155 G; 343 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 25; DB 11; Length 1085;
XX Best Local Similarity 100.0%; Pred. No. 0.007;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 ggatcccccgatgagctaaagctagc 25
Db 815 ggatcccccgatgagctaaagctagc 839
|||||
XX
RESULT 5
AAQ04706
ID AAQ04706 standard; DNA; 1160 BP.
XX
XX
XX AC AAQ04706;
XX
XX 12-OCT-1990 (first entry)
XX
XX USP-signalpeptide cassette USP-sig.T7.
XX
XX Foreign DNA incorporation; recombinant DNA techniques;
KW higher plant genome; signalpeptide; USP-Sig.T7.; ss.
FH Key Location/Qualifiers
FT CDS 708..877
FT /*tag= a
FT /product=signalpeptide
FT 747..817
XX
XX DE3920034-A.
XX
XX 31-MAY-1990.
XX
XX 20-JUN-1989; 89DE-3920034.
XX
XX 19-SEP-1988; 88DD-0319887.
XX
```

```
XX (PFLA-) VE KOMB PFLANZENZUC.
XX
XX Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;
XX WPI; 1990-172459/23.
XX P-PSDB; AAR05199.
XX
XX Incorporation of DNA into higher plant genome - by specified
XX recombinant DNA techniques.
XX
XX Disclosure; ; pp; German.
XX
XX The unique BglII-Ort (890-895) site is for
XX ligating foreign DNA and the HindIII-Ort in the 3' polylinker
XX (1155-1160) for cloning the cassette in the Ti-vector pGA471.
XX The cassette is cloned into the binary Ti-vectors pGA471 and
XX Agrobacterium tumefaciens is transfected.
XX See also AAQ04703-Q04706.
XX
XX Sequence 1160 BP; 369 A; 192 C; 188 G; 411 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 25; DB 11; Length 1160;
XX Best Local Similarity 100.0%; Pred. No. 0.007;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 ggatcccccgatgagctaaagctagc 25
Db 890 ggatcccccgatgagctaaagctagc 914
|||||
XX
RESULT 6
AAT39337
ID AAT39337 standard; DNA; 1303 BP.
XX
XX AC AAT39337;
XX
XX 22-JAN-1997 (first entry)
XX
XX Plasmid pTS88 (EcoRI-HindIII fragment).
XX
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH misc_feature 1..35
FT /*tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT 36..694
FT /*tag= b
FT /label= p35S
FT /function= 35S promoter of cauliflower mosaic virus
FT strain CM1841
XX
XX CDS 695..967
XX /*tag= c
XX /label= barstar
XX /product= Bacillus amyloliquefaciens barstar
XX 968..1287
XX /*tag= d
XX /label= 3'g7
XX /function= region containing polyadenylation signal
XX of gene 7 og Agrobacterium T-DNA
XX 1288..1303
XX /*tag= e
XX /label= pGEM2
XX /note= "polylinker of pGEM2"
XX
XX WO9626283-A1.
XX
```

```

PD 29-AUG-1996.
XX
PF 21-FEB-1996; 96WO-EP00722.
XX
PR 21-FEB-1995; 95EP-0400364.
XX
PA (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
PI Botterman J, Cornelissen M, Michiels F;
XX WPI; 1996-402373/40.
XX
DR Prodn. of male sterile plants by transforming with a chimaeric
PT construct - comprising a male sterility DNA e.g. barnase and a
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
PT for generating hybrid cultivars
XX
XX Example 1; Page 38; 56pp; English.
XX
CC The HindIII-EcoRI fragment (AAT39337) of plasmid pT888 contains
CC barstar DNA under control of a 35S promoter. The plasmid was
CC used with pT8174 (see also AAT39336) contg. barnase DNA under
CC control of the stamen-specific promoter El to produce male sterile
CC rice cv. Kochihibiki transgenic plants, and with plasmid pVE136
CC (see also AAT39338) contg. barnase DNA under control of the stamen-
CC specific PCA35 promoter to produce male sterile maize plants.
CC Expression of barnase (a ribonuclease) in the stamen leads to male
CC sterility. Constitutive expression of barstar counteracts possible
CC low level expression of barnase DNA in non-stamen tissue.
XX
SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 25; DB 17; Length 1303;
Best Local Similarity 100.0%; Pred. No. 0.0071;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggatcccccgcgtagcgaagctagc 25
Db 1024 ggaatcccccgcgtagcgaagctagc 1048

RESULT 7
AAH25423/c
ID AAH25423 standard; DNA; 4832 BP.
XX
AC AAH25423;
XX
XX 22-AUG-2001 (first entry)
XX
DE Nucleotide sequence of plasmid pTHW118.
XX
KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barstar gene; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
OS Arabidopsis thaliana.
OS Bacillus amyloliquefaciens.
OS Nicotiana tabacum.
XX
XX Key Location/Qualifiers
FH misc_feature 1..25
FT /*tag= a
FT /note= "right border repeat from TL-DNA from pTiB6S3"
FT 26..53
FT /*tag= b
FT /note= "synthetic polylinker derived sequences"
FT 54..90
FT /*tag= c
FT /note= "residual sequence from TL-DNA at right
FT border repeat"
FT 91..97

```

```

FT FT /*tag= d
FT FT /note= "synthetic polylinker derived sequences"
FT FT complement (98..309)
FT FT /*tag= e
FT FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"
FT FT 310..330
FT FT /*tag= f
FT FT /note= "synthetic polylinker derived sequences"
FT FT complement (331..882)
FT FT /*tag= g
FT FT /note= "Streptomyces hygroscopicus bialaphos
FT FT resistance (bar) gene"
FT FT complement (883..2608)
FT FT /*tag= h
FT FT /note= "atsIA ribulose-1,5-biphosphate carboxylase
FT FT small subunit gene from Arabidopsis thaliana"
FT FT 2609..2658
FT FT /*tag= i
FT FT /note= "synthetic polylinker derived sequences"
FT FT complement (2659..2919)
FT FT /*tag= j
FT FT /note= "TaqI fragment from 3' UTR of nopaline
FT FT synthase gene from T-DNA of pTiT37 and
FT FT containing plant polyadenylation signals"
FT FT 2920..2940
FT FT /*tag= k
FT FT /note= "synthetic polylinker derived sequences"
FT FT 2941..2980
FT FT /*tag= l
FT FT /note= "downstream of Bacillus amyloliquefaciens
FT FT barstar coding region"
FT FT complement (2981..3253)
FT FT /*tag= m
FT FT /note= "Barstar gene coding region from Bacillus
FT FT amyloliquefaciens"
FT FT complement (3254..4762)
FT FT /*tag= n
FT FT /note= "anther-specific gene TA29 promoter from
FT FT Nicotiana tabacum"
FT FT 4763..4807
FT FT /*tag= o
FT FT /note= "synthetic polylinker derived sequences"
FT FT 4808..4832
FT FT /*tag= p
FT FT /note= "left border repeat from TL-DNA from pTiB6S3"
XX
XX WO200141558-A1.
XX
XX 14-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-BP12872.
XX
XX 08-DEC-1999; 99US-0457037.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX De Both G, De Beuckeleer M;
XX WPI; 2001-381419/40.
XX
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
XX with improved qualities, comprises a male-sterility gene and fertility
XX restorer gene, integrated into the genome -
XX
XX Example 1; Page 80-82; 98pp; English.
XX
XX The specification describes a pair of transgenic winter oilseed rape
XX plants suited for producing hybrid seed. One of the plants has an
XX expression cassette comprising a male-sterility gene, and the other
XX plant has an expression cassette comprising a fertility restorer gene,
XX integrated into the genome. The fertility restorer gene is capable of
XX preventing the activity of the male-sterility gene. The plant pair is
XX useful for producing hybrid seed. Plants developed from the hybrid

```

```

CC seed have agronomic performance, genetic stability and adaptability to
CC different genetic backgrounds. The present sequence represents
CC plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
CC a fertility restorer gene. The plasmid is used to create transgenic
CC plants of the invention.
XX
SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match      100.0%; Score 25; DB 22; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.0078;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaagctagc 25
   |||||
Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

RESULT 8
AAT59531/c
ID AAT59531 standard; DNA; 4946 BP.
XX
AC AAT59531;
XX
DT 07-MAY-1997 (first entry)
XX
DE T-DNA of plasmid pTHW107.
XX
KW Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PAPP;
KW niacinamide; Agrobacterium; T-DNA; male sterile; barnase;
KW ribonuclease; Rns; cereal; wheat; oilseed rape; Brassica napus;
KW plasmid pTHW107; ds.
XX
OS Chimeric Agrobacterium sp.;
OS Chimeric Arabidopsis thaliana;
OS Chimeric Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT misc_RNA complement (1..25)
FT /tag= a
FT /label= RB
FT /note= "T-DNA right border"
FT complement (97..330)
FT /tag= b
FT /label= 3'g7
FT /note= "3' untranslated region contg. the poly-A
FT signal of gene-7 of Agrobacterium T-DNA"
FT complement (331..882)
FT /tag= c
FT /label= Bar
FT /product= phosphinothricin acetyltransferase
FT complement (883..2608)
FT /tag= d
FT /label= PSSU
FT /note= "promoter region of Rubisco small subunit
FT gene of Arabidopsis thaliana"
FT complement (2658..3031)
FT /tag= e
FT /label= 3'nos
FT /note= "3' untranslated region contg. the poly-A
FT signal of the nopaline-synthase gene of
FT Agrobacterium T-DNA"
FT complement (3032..3367)
FT /tag= f
FT /label= Barnase
FT /product= barnase
FT complement (3368..4876)
FT /tag= g
FT /label= PTA29
FT /note= "promoter region of tobacco TA29 gene"
FT complement (4822..4946)
FT /tag= h
FT /label= LB
FT

```

```

FT /note= "T-DNA left border"
XX
PN EP757102-A1.
XX
PD 05-FEB-1997.
XX
PF 04-AUG-1995; 95EP-0401844.
XX
PR 04-AUG-1995; 95EP-0401844.
XX
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
PI De Block M;
XX
DR WPI; 1997-1111050/11.
XX
PT Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
PT inhibitor - reduces the cultured cells response to stress and
PT reduces metabolism
XX
PS Example 3; Page 13-16; 25pp; English.
XX
CC Plasmid pTHW107 is a vector carrying T-DNA (AAT59531) comprising a
CC barnase coding sequence under control of the tobacco TA29 gene
CC stamen-specific promoter and a phosphinothricin acetyltransferase
CC coding sequence under control of an Arabidopsis Rubisco small
CC subunit gene promoter. Oilseed rape hypocotyl explants were
CC infected with Agrobacterium tumefaciens C58C1Rif carrying vector
CC pTHW107 and helper Ti plasmid pMP60. In some treatments, the
CC hypocotyls were treated with the poly-(ADP-ribose) polymerase
CC inhibitor niacinamide (250 mg/l) 4 days prior to infection.
CC Plants regenerated from niacinamide-treated transformed calli
CC had a low copy number and displayed less variation in the
CC expression profile of the transgenes.
XX
SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match      100.0%; Score 25; DB 18; Length 4946;
Best Local Similarity 100.0%; Pred. No. 0.0078;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatccccgatgagctaagctagc 25
   |||||
Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

RESULT 9
AAH25422/c
ID AAH25422 standard; DNA; 4946 BP.
XX
AC AAH25422;
XX
DT 22-AUG-2001 (first entry)
XX
DE Nucleotide sequence of plasmid pTHW107.
XX
KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW fertility restorer gene; barnase gene; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
OS Arabidopsis thaliana.
OS Bacillus amyloliquefaciens.
OS Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "right border repeat from TL-DNA from pTHB6S3"
FT misc_feature 26..97
FT /tag= b
FT /note= "synthetic polylinker derived sequences"
FT

```

FT 3'UTR complement (98..309)  
FT /\*tag= c  
FT /note= "3' UTR from TL-DNA gene 7 of pT1B6S3"  
FT 310..330  
FT /\*tag= d  
FT /note= "synthetic polylinker derived sequences"  
FT 331..882  
FT /\*tag= e  
FT /note= "Streptomyces hygrosopicus bar gene"  
FT complement (883..2608)  
FT /\*tag= f  
FT /note= "ats1A ribulose-1,5-biphosphate carboxylase  
FT small subunit gene from Arabidopsis thaliana"  
FT 2609..2658  
FT /\*tag= g  
FT /note= "synthetic polylinker derived sequences"  
FT complement (2659..2919)  
FT /\*tag= h  
FT /note= "taqI fragment from 3' UTR of nopaline  
FT synthase gene from T-DNA of pT137 and  
FT containing plant polyadenylation signals"  
FT 2920..3031  
FT /\*tag= i  
FT /note= "3'UTR downstream of Bacillus amyloliquefaciens  
FT barnase coding region"  
FT complement (3032..3367)  
FT /\*tag= j  
FT /note= "Barnase coding region from Bacillus  
FT amyloliquefaciens"  
FT complement (3368..4877)  
FT /\*tag= k  
FT /note= "anther-specific gene TA29 promoter from  
FT Nicotiana tabacum"  
FT 4878..4921  
FT /\*tag= l  
FT /note= "synthetic polylinker derived sequences"  
FT 4922..4946  
FT /\*tag= m  
FT /note= "left border repeat from TL-DNA from pT1B6S3"  
FT WO200141558-A1.  
XX 14-JUN-2001.  
XX  
XX  
XX 06-DEC-2000; 2000WO-EPI2872.  
XX  
XX 08-DEC-1999; 99US-0457037.  
XX (AVET ) AVENTIS CROPSCEINCE NV.  
XX  
XX De Both G, De Beuckeleer M;  
XX WPI; 2001-381419/40.  
XX  
XX Transgenic winter oilseed rape plants suited for producing hybrid seed  
XX with improved qualities, comprises a male-sterility gene and fertility  
XX restorer gene, integrated into the genome  
XX  
XX Example 1; Page 78-80; 98pp; English.  
XX  
XX The specification describes a pair of transgenic winter oilseed rape  
XX plants suited for producing hybrid seed. One of the plants has an  
XX expression cassette comprising a male-sterility gene, and the other  
XX plant has an expression cassette comprising a fertility restorer gene,  
XX integrated into the genome. The fertility restorer gene is capable of  
XX preventing the activity of the male-sterility gene. The plant pair is  
XX useful for producing hybrid seed. Plants developed from the hybrid  
XX seed have agronomic performance, genetic stability and adaptability to  
XX different genetic backgrounds. The present sequence represents  
XX a plasmid pT1W107. This plasmid comprises the barnase gene, which acts as  
XX a male-sterility gene. The plasmid is used to create transgenic plants  
XX of the invention.

SQ Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;  
  
Query Match 100.0%; Score 25; DB 22; Length 4946;  
Best Local Similarity 100.0%; Pred. No. 0.0078;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ggateccccgatgagctaaagctagc 25  
Db 317 GGATCCCCCGATGAGCTAAGCTAGC 293  
|||||  
  
RESULT 10  
AAF86439  
ID AAF86439 standard; DNA; 5228 BP.  
XX  
XX AAF86439;  
XX AC  
XX XX  
XX 25-JUN-2001 (first entry)  
XX XX  
XX Plasmid pT1S172delta.  
XX DE  
XX Male sterile plant; RNAase inhibitor; plasmid pT1S172delta; ds.  
XX KW  
XX Unidentified.  
XX OS  
XX WO200124616-A1.  
XX PN  
XX 12-APR-2001.  
XX PD  
XX 12-SEP-2000; 2000WO-JP06222.  
XX PF  
XX 30-SEP-1999; 99JP-0279307.  
XX PR  
XX (NISB ) JAPAN TOBACCO INC.  
XX PA  
XX Hamada K, Nakakido F;  
XX PI  
XX WPI; 2001-266212/27.  
XX DR  
XX  
XX Method for producing male sterile rice and maize by inserting RNAse  
XX gene and RNAse inhibitor genes with promoters into the plant genome -  
XX  
XX Disclosure; Page 14-17; 29pp; Japanese.  
XX  
XX The present invention relates to a method for producing male sterile  
XX plants. The method comprises inserting a promoter fragment upstream of an  
XX RNAse gene and a second promoter, upstream of an RNAse inhibitor protein  
XX gene and inserting it into the plant genome. The method is useful for  
XX producing male sterile tobacco, lettuce and rapeseed plants, but  
XX preferably rice and maize. The present sequence is a vector used in  
XX the method of the present invention.  
XX  
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;  
  
Query Match 100.0%; Score 25; DB 22; Length 5228;  
Best Local Similarity 100.0%; Pred. No. 0.0079;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ggateccccgatgagctaaagctagc 25  
Db 4957 ggatcccccgatgagctaaagctagc 4981  
|||||  
  
RESULT 11  
AAV23239/C  
ID AAV23239 standard; DNA; 5349 BP.  
XX  
XX AAV23239;  
XX AC  
XX 17-JUL-1998 (first entry)  
XX DT  
XX

```

DE T-DNA of pTTS24.
XX
KW Barstar; barnase inhibitor; fertility restoration;
KW male-sterile line; plasmid pTTS24; T-DNA; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature complement (1..25)
FT /*tag= a
FT /label= RB
FT /note= "right boarder"
FT /*tag= b
FT /label= 3' g7
FT /note= "region containing 3' untranslated end of
FT Agrobacterium T-DNA gene 7"
FT
FT CDS 332..883
FT /*tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin acetyl
FT transferase"
FT complement (884..2258)
FT /*tag= d
FT /label= P35S
FT /note= "35S promoter of Cauliflower Mosaic Virus"
FT 2281..3969
FT /*tag= e
FT /label= PE1
FT /note= "promoter of E1 gene of rice (W09213956)"
FT 3970..4245
FT /*tag= f
FT /product= improved_barstar
FT 4246..4577
FT /*tag= g
FT /label= 3'_chs
FT /note= "region containing 3' untranslated end of
FT chalcone synthase gene"
FT complement (5325..5349)
FT /*tag= h
FT /note= "T-DNA left border"
XX
PN W09810081-A2.
XX
XX 12-MAR-1998.
XX
XX 01-SEP-1997; 97WO-EP04739.
XX
XX 03-SEP-1996; 96EP-0202446.
XX
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX Michiels F, Williams M;
XX
XX WPI; 1998-193630/17.
XX
XX DNA encoding an improved barstar protein - used to restore fertility
XX in male-sterile plant lines
XX
XX Example 4; Pages 41-43; 54pp; English.
XX
XX The present sequence was used in the preparation of an improved
XX Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which
XX can be used to restore fertility to male-sterile lines.
XX The DNA sequence encoding the improved barstar, leads to increased
XX barstar production in tapetum cells, due to improved translation,
XX and possibly protein stability.
XX
XX Sequence 5349 BP; 1339 A; 1233 C; 1290 G; 1487 T; 0 other;
XX
XX Query Match 100.0%; Score 25; DB 19; Length 5349;
XX Best Local Similarity 100.0%; Pred. No. 0.0079;

```

```

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatcccccagatgacgctaacgtacg 25
   |||||
DB 317 GGAATCCCGGATGAGCTAAGCTAGC 293

RESULT 12
AAT39339
ID AAT39339 standard; DNA; 5864 BP.
XX
AC AAT39339;
XX
DT 22-JAN-1997 (first entry)
XX
DE Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.
XX
KW Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;
KW transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature complement (1..25)
FT /*tag= a
FT /label= RB
FT /note= "right border of Agrobacterium T-DNA"
FT complement (98..330)
FT /*tag= b
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT complement (331..882)
FT /*tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin
FT acetyltransferase"
FT complement (883..2608)
FT /*tag= d
FT /label= PSSu
FT /note= "promoter of Arabidopsis Rubisco small
FT subunit gene"
FT complement (2659..3031)
FT /*tag= e
FT /label= 3'nos
FT /note= "region containing polyA signal of nopaline
FT synthase gene of Agrobacterium T-DNA"
FT complement (3032..3367)
FT /*tag= f
FT /label= Barnase
FT /note= "Bacillus amyloliquefaciens barnase coding
FT region"
FT complement (3368..4877)
FT /*tag= g
FT /label= pTA29
FT /note= "promoter of stamen-specific TA29 gene of
FT Nicotiana tabacum"
FT 4924..5216
FT /*tag= h
FT /label= Pnos
FT /note= "promoter of nopaline synthase gene of
FT Agrobacterium T-DNA"
FT 5217..5489
FT /*tag= i
FT /label= Barstar
FT /note= "region coding for barstar of Bacillus
FT amyloliquefaciens"
FT 5490..5765
FT /*tag= j
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT complement (5840..5864)
FT misc_feature

```

```

FT      /*tag= k
FT      /label= LB
FT      /note= "left border of Agrobacterium T-DNA"
PN      WO9626283-A1.
XX
XX      29-AUG-1996.
XX
XX      21-FEB-1996; 96WO-EP00722.
XX
XX      21-FEB-1995; 95EP-0400364.
XX
XX      (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX      Botterman J, Cornelissen M, Michiels F;
XX      WPI; 1996-402373/40.
XX
XX      Prodn. of male sterile plants by transforming with a chimaeric
XX      construct - comprising a male sterility DNA e.g. barnase and a
XX      co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX      for generating hybrid cultivars
XX
XX      Example 3; Page 33-3743-47; 56pp; English.
XX
XX      Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
XX      under control of the PSU promoter, a barnase gene under control
XX      of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
XX      gene under control of the Pnos promoter. 87% Of oilseed rape
XX      plants regenerated after Agrobacterium-mediated transformation
XX      using pTCO113 were male sterile. Barnase expression disturbed the
XX      function of stamen cells leading to male sterility. Constitutive
XX      expression of barstar counteracted any low level expression of
XX      barnase in non-stamen tissue.
XX
XX      Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match      100.0%; Score 25; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ggatccccgatgagctaaagctagc 25
Db      |||||
5546 ggatccccgatgagctaaagctagc 5570

RESULT 13
AAT39339/C
ID      AAT39339 standard; DNA; 5864 BP.
XX
XX      AAT39339;
XX
XX      22-JAN-1997 (first entry)
XX
XX      Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.
XX
XX      Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;
XX      transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX      Synthetic.
XX
XX      Key      Location/Qualifiers
XX      misc_feature complement (1..25)
XX      /*tag= a
XX      /label= RB
XX      /note= "right border of Agrobacterium T-DNA"
XX      complement (98...330)
XX      /*tag= b
XX      /label= 3'g7
XX      /note= "region containing polyA signal of gene 7
XX      of Agrobacterium T-DNA"
XX      complement (331..882)
XX      CDS

```

```

FT      /*tag= c
FT      /label= bar
FT      /note= "region coding for phosphinothricin
FT      acetyltransferase"
FT      complement (883..2608)
FT      /*tag= d
FT      /label= Psu
FT      /note= "promoter of Arabidopsis Rubisco small
FT      subunit gene"
FT      complement (2659..3031)
FT      /*tag= e
FT      /label= 3'nos
FT      /note= "region containing polyA signal of nopaline
FT      synthase gene of Agrobacterium T-DNA"
FT      complement (3032..3367)
FT      /*tag= f
FT      /label= Barnase
FT      /note= "Bacillus amyloliquefaciens barnase coding
FT      region"
FT      complement (3368..4877)
FT      /*tag= g
FT      /label= PTA29
FT      /note= "promoter of stamen-specific TA29 gene of
FT      Nicotiana tabacum"
FT      4924..5216
FT      /*tag= h
FT      /label= Pnos
FT      /note= "promoter of nopaline synthase gene of
FT      Agrobacterium T-DNA"
FT      5217..5489
FT      /*tag= i
FT      /label= Barstar
FT      /note= "region coding for barstar of Bacillus
FT      amyloliquefaciens"
FT      5490..5765
FT      /*tag= j
FT      /label= 3'g7
FT      /note= "region containing polyA signal of gene 7
FT      of Agrobacterium T-DNA"
FT      complement (5840..5864)
FT      /*tag= k
FT      /label= LB
FT      /note= "left border of Agrobacterium T-DNA"
XX
XX      WO9626283-A1.
XX
XX      29-AUG-1996.
XX
XX      21-FEB-1996; 96WO-EP00722.
XX
XX      21-FEB-1995; 95EP-0400364.
XX
XX      (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX      Botterman J, Cornelissen M, Michiels F;
XX      WPI; 1996-402373/40.
XX
XX      Prodn. of male sterile plants by transforming with a chimaeric
XX      construct - comprising a male sterility DNA e.g. barnase and a
XX      co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX      for generating hybrid cultivars
XX
XX      Example 3; Page 33-3743-47; 56pp; English.
XX
XX      Plasmid pTCO113 (AAT39339) is a T-DNA vector containing a bar gene
XX      under control of the PSU promoter, a barnase gene under control
XX      of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
XX      gene under control of the Pnos promoter. 87% Of oilseed rape
XX      plants regenerated after Agrobacterium-mediated transformation
XX      using pTCO113 were male sterile. Barnase expression disturbed the
XX      function of stamen cells leading to male sterility. Constitutive
XX      expression of barstar counteracted any low level expression of

```



```

CC barnase in non-stamen tissue.
XX Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;
SQ

Query Match      100.0%; Score 25; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgactaagctagc 25
Db 317 GGATCCCCCGATGACTAGCTAGC 293

RESULT 14
AAD06990
ID AAD06990 standard; DNA; 5865 BP.
XX
AC AAD06990;
XX
DT 06-AUG-2001 (first entry)
XX
DE Chimeric T-DNA of plasmid pTC0113.
XX
KW T-DNA; plasmid pTC0113; transgenic Brassica plant; transformation event;
KW male-sterility gene; chimeric; tobacco; ds.
XX
OS Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "Right border repeat from the TL-DNA from
FT pTiB6S3"
FT misc_feature 26..53
FT /tag= b
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 54..90
FT /tag= c
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT 3'UTR 98..309
FT /tag= d
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 310..331
FT /tag= e
FT /note= "Synthetic polylinker derived sequence"
FT CDS 332..883
FT /tag= f
FT /product= "Protein encoded by bialaphos resistance
FT gene (bar) of Streptomyces hygroscopicus"
FT promoter 884..2609
FT /tag= g
FT /note= "Promoter from the atSLA ribulose-1,5-bisphosphate
FT carboxylase small subunit gene from Arabidopsis thaliana"
FT misc_feature 2610..2659
FT /tag= h
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 2660..2920
FT /tag= i
FT /note= "Tag1 fragment from the 3' untranslated end of the
FT nopaline synthase gene (3'nos) from the T-DNA of pTiT37
FT and containing plant polyadenylation signals"
FT misc_feature 2921..2936
FT /tag= j
FT /note= "Synthetic polylinker derived sequence"
FT 3'UTR 2937..3032

```

---

```

FT /tag= k
FT /note= "The 3' untranslated region downstream from the
FT barnase coding sequence of Bacillus amyloliquefaciens"
FT CDS 3033..3368
FT /tag= l
FT /product= "Protein encoded by barnase gene from
FT Bacillus amyloliquefaciens"
FT promoter 3369..4878
FT /tag= m
FT /note= "Promoter region of the anther-specific gene
FT TA29 from Nicotiana tabacum"
FT misc_feature 4879..4924
FT /tag= n
FT /note= "Synthetic polylinker derived sequence"
FT 4925..5215
FT /tag= o
FT /note= "Promoter of the nopaline synthase gene from the
FT T-DNA of pTiT37 of Agrobacterium tumefaciens"
FT misc_feature 5216..5217
FT /tag= p
FT /note= "Synthetic polylinker derived sequence"
FT CDS 5218..5490
FT /tag= q
FT /product= "Protein encoded by barstar gene of
FT Bacillus amyloliquefaciens"
FT misc_feature 5491..5530
FT /tag= r
FT /note= "Sequence from the 3' untranslated end of the
FT barstar gene from Bacillus amyloliquefaciens"
FT misc_feature 5531..5554
FT /tag= s
FT /note= "Synthetic polylinker derived sequence"
FT 5555..5766
FT /tag= t
FT /note= "The 3' untranslated end from the TL-DNA
FT gene 7 (3'g7) of pTiB6S3"
FT misc_feature 5767..5773
FT /tag= u
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 5774..5810
FT /tag= v
FT /note= "Residual sequence from the TL-DNA at the
FT right border repeat"
FT misc_feature 5811..5840
FT /tag= w
FT /note= "Synthetic polylinker derived sequence"
FT misc_feature 5841..5865
FT /tag= x
FT /note= "Left border repeat from the TL-DNA from
FT pTiB6S3"
XX WO200131042-A2.
XX
PD 03-MAY-2001.
XX
PF 26-OCT-2000; 2000WO-EP10680.
XX
PR 29-OCT-1999; 99US-0430497.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX Weston B, De Beuckeleer M;
XX WPI; 2001-300517/31.
XX
PT Transgenic Brassica plants, seeds, cells or tissues, characterized by
PT harboring specific transformation events, particularly by presence of
PT male-sterility gene, at specific location in its genome.
XX
XX Claim 1; Page 47-49; 53pp; English.
XX
CC The present invention relates to a transgenic Brassica plant or its
CC seed, cells or tissues, characterised by harbouring a specific

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:13 ; Search time 301.6 Seconds  
(without alignments)  
18.773 Million cell updates/sec

Title: US-09-698-903B-5

Perfect score: 25  
Sequence: 1 ggatccccgatgactaagctagc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1303	US-08-894-440-2	Sequence 2, Appli
c 2	25	100.0	4946	US-08-817-188-1	Sequence 1, Appli
c 3	25	100.0	5560	US-08-817-188-5	Sequence 5, Appli
4	25	100.0	5864	US-08-894-440-4	Sequence 4, Appli
c 5	25	100.0	5864	US-08-894-440-4	Sequence 4, Appli
6	25	100.0	6548	US-08-894-440-1	Sequence 1, Appli
7	25	100.0	6548	US-08-817-188-2	Sequence 2, Appli
8	24	96.0	3200	US-08-453-104-23	Sequence 23, Appli
9	24	96.0	3200	US-08-694-824-23	Sequence 23, Appli
10	24	96.0	3201	US-08-453-104-22	Sequence 22, Appli
11	24	96.0	3201	US-08-694-824-22	Sequence 22, Appli
12	24	96.0	7566	US-08-232-016-23	Sequence 23, Appli
13	24	96.0	7639	US-08-232-016-22	Sequence 22, Appli
c 14	24	96.0	7811	US-08-549-680A-5	Sequence 5, Appli
c 15	17	68.0	29	US-08-232-016-11	Sequence 11, Appli
16	17	68.0	3153	US-09-080-625-3	Sequence 3, Appli
17	17	68.0	3336	US-09-080-625-2	Sequence 2, Appli
18	17	68.0	3694	US-09-080-625-5	Sequence 5, Appli
19	17	68.0	3877	US-09-080-625-4	Sequence 4, Appli
c 20	17	68.0	24595	5428147-1	Patent No. 5428147
c 21	16.6	66.4	6201	US-08-790-912-1	Sequence 1, Appli
22	16	64.0	54	US-08-503-226B-32	Sequence 32, Appli
23	16	64.0	54	US-08-721-458B-32	Sequence 32, Appli
c 24	16	64.0	3065	US-08-209-521-28	Sequence 28, Appli
25	15.4	61.6	1186	US-08-084-121-2	Sequence 2, Appli
26	15.4	61.6	1186	US-08-478-015-2	Sequence 2, Appli
27	15.4	61.6	1186	US-08-475-975-2	Sequence 2, Appli

28	15.4	61.6	1186	3	US-09-084-889-2	Sequence 2, Appli
29	15.4	61.6	1859	3	US-08-691-563C-46	Sequence 46, Appli
c 30	15.4	61.6	2143	2	US-08-656-177A-1	Sequence 1, Appli
c 31	15.4	61.6	2143	3	US-09-256-797-1	Sequence 1, Appli
c 32	15.4	61.6	10281	2	US-08-816-155B-1	Sequence 1, Appli
c 33	15.4	61.6	10281	3	US-09-079-587-1	Sequence 1, Appli
c 34	15.2	60.8	43795	3	US-08-742-185-101	Sequence 101, App
35	15	60.0	1642	1	US-07-996-772A-1	Sequence 1, Appli
36	15	60.0	1642	1	US-08-446-822-1	Sequence 1, Appli
37	15	60.0	1642	5	PCT-US93-12586-1	Sequence 1, Appli
c 38	15	60.0	2352	2	US-08-922-837-1	Sequence 1, Appli
c 39	15	60.0	2352	4	US-09-351-550-1	Sequence 1, Appli
c 40	15	60.0	7431	4	US-09-306-998-2	Sequence 2, Appli
c 41	14.8	59.2	1377	1	US-08-112-817C-1	Sequence 1, Appli
42	14.8	59.2	1710	4	US-09-068-960-14	Sequence 14, Appli
43	14.6	58.4	376	3	US-08-985-950-3	Sequence 3, Appli
c 44	14.6	58.4	699	4	US-08-998-416-591	Sequence 591, App
c 45	14.4	57.6	60	4	US-09-171-945-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1  
US-08-894-440-2  
; Sequence 2, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894,440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI  
; OTHER INFORMATION: fragment of pT588  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (p35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amyloliquefactions  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
; US-08-894-440-2

Query Match 100.0%; Score 25; DB 3; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.0011;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ggatccccgatgactaagctagc 25

Db 1024 ggatcccccgatgagctaagctagc 1048  
 |||

RESULT 2  
 US-08-817-188-1/c  
 ; Sequence 1, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
 ; EARLIER FILING DATE: 1996-07-31  
 ; EARLIER APPLICATION NUMBER: EP 95401844.6  
 ; EARLIER FILING DATE: 1995-08-04  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 4946  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
 ; OTHER INFORMATION: plasmid pTHW107  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement(1)..(25)  
 ; OTHER INFORMATION: T-DNA right border (RB)  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((97)..(330))  
 ; OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((331)..(882))  
 ; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
 ; OTHER INFORMATION: transferase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((883)..(2608))  
 ; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
 ; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((2658)..(3031))  
 ; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3032)..(3367))  
 ; OTHER INFORMATION: barnase: region coding for barnase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3368)..(4876))  
 ; OTHER INFORMATION: pTA29: promoter region of TA29 gene of Nicotiana  
 ; OTHER INFORMATION: tabacum  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((4922)..(4946))  
 ; OTHER INFORMATION: LB: T-DNA left border  
 US-08-817-188-1

Query Match 100.08; Score 25; DB 3; Length 4946;  
 Best Local Similarity 100.08; Pred. No. 0.0014;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaagctagc 25  
 |||  
 Db 317 GGATCCCCGATGAGCTAAGCTAGC 293

RESULT 3  
 US-08-817-188-5/c  
 ; Sequence 5, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
 ; EARLIER FILING DATE: 1996-07-31  
 ; EARLIER APPLICATION NUMBER: EP 95401844.6  
 ; EARLIER FILING DATE: 1995-08-04  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 5560  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
 ; OTHER INFORMATION: plasmid pTHW142  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(25)  
 ; OTHER INFORMATION: pTIB6S3  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (84)..(296)  
 ; OTHER INFORMATION: 3' g7: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (318)..(869)  
 ; OTHER INFORMATION: bar: region coding for phosphinotricin  
 ; OTHER INFORMATION: acetyltransferase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (830)..(2760)  
 ; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit  
 ; OTHER INFORMATION: gene of Arabidopsis thaliana  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (2765)..(3058)  
 ; OTHER INFORMATION: 3' untranslated region of the CaMV 35S transcript  
 ; OTHER INFORMATION: containing polyadenylation signals  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (3059)..(5056)  
 ; OTHER INFORMATION: uidA: region coding for beta-glucuronidase  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4483)..(4671)  
 ; OTHER INFORMATION: IV2: region corresponding to the second intron of  
 ; OTHER INFORMATION: the ST-LS1 gene  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (5087)..(5502)  
 ; OTHER INFORMATION: P35S: 35S promoter region of CaMV  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (5533)..(5560)  
 ; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from  
 ; OTHER INFORMATION: pTIB6S3  
 ; FEATURE:

```
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; US-08-817-188-5

Query Match      100.0%; Score 25; DB 3; Length 5560;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25
   |||||
Db 304 GGATCCCCCGATGAGCTAAGCTAGC 280

RESULT 4
US-08-894-440-4
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTC0113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
```

```
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Phos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
; US-08-894-440-4

Query Match      100.0%; Score 25; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25
   |||||
Db 5546 ggatcccccgatgagctaaagctagc 5570

RESULT 5
US-08-894-440-4/c
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTC0113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

```

```

Query Match      100.0%; Score 25; DB 3; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
    ||||||||||||||||||||||||
Db 317 GGATCCCCCGATGAGCTAGCTAGC 293

```

```

RESULT 6
US-08-894-440-1
; Sequence 1, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTS174
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2003)
; OTHER INFORMATION: pUC19 derived vector sequences (vector)
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2283))
; OTHER INFORMATION: 3' nos: region containing polyadenylation signal
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2284)..(2624))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: promoter of the stamen-specific B1 gene of rice
; OTHER INFORMATION: (PEI)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5730)
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus (P35S)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5711)..(6262)
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: region containing polyadenylation signal fo gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
US-08-894-440-1

```

```

Query Match      100.0%; Score 25; DB 3; Length 6548;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcccccgatgagctaaagctagc 25
    ||||||||||||||||||||||||
Db 6277 ggatcccccgatgagctaaagctagc 6301

```

```

RESULT 7
US-08-817-188-2
; Sequence 2, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTS172
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2288))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2289)..(2624))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:

```



```
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: PE1: promoter region of E1 gene of rice
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5170)
; OTHER INFORMATION: P35S: 35S promoter region of Cauliflower Mosaic
; OTHER INFORMATION: Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5711)..(6262))
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
US-08-817-188-2
```

Query Match 100.0%; Score 25; DB 3; Length 6548;  
Best Local Similarity 100.0%; Pred. No. 0.0014;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatcccccgatgagctaaagctagc 25  
|||||

Db 6277 ggatcccccgatgagctaaagctagc 6301

```
RESULT 8
US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
US-08-453-104-23

Query Match 96.0%; Score 24; DB 1; Length 3200;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gatcccccgatgagctaaagctagc 25
|||||

Db 2879 GATCCCCGATGAGCTAAGCTAGC 2902

RESULT 9
US-08-694-824-23
; Sequence 23, Application US/08694824
; Patent No. 5877306
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
```

; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 96.0%; Score 24; DB 2; Length 3200;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gatccccgatgactaagctagc 25  
|||||  
Db 2879 GATCCCCGATGACTAAGCTAGC 2902

RESULT 10  
US-08-453-104-22  
; Sequence 22, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,104  
FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,869  
FILING DATE: 16-DEC-1992  
APPLICATION NUMBER: GB 90401055.0  
FILING DATE: 18-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Rea, Teresa S  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 010830-032  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3201 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 2151..2155  
OTHER INFORMATION: /note= "Nucleotides 2151-2155  
OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 96.0%; Score 24; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.004;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 gatccccgatgactaagctagc 25  
|||||  
Db 2952 GATCCCCGATGACTAAGCTAGC 2975

RESULT 11  
US-08-694-824-22  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/694,824  
FILING DATE: 09-AUG-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,869  
FILING DATE: 16-DEC-1992  
APPLICATION NUMBER: GB 90401055.0  
FILING DATE: 18-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Rea, Teresa S  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 010830-032  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3201 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:

NAME/KEY: misc feature  
LOCATION: 2151..2155  
OTHER INFORMATION: /note= "Nucleotides 2151-2155  
OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 96.0%; Score 24; DB 2; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gatccccgatgactaagctagc 25  
|||||  
Db 2952 GATCCCCGATGACTAAGCTAGC 2975

RESULT 12

US-08-232-016-23  
; Sequence 23, Application US/08232016  
; Patent No. 5952547  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; APPLICANT: VAN AARSEN, Roel  
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
; TITLE OF INVENTION: PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.016  
; FILING DATE: 03-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 91402920.2  
; FILING DATE: 30-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 92400820.4  
; FILING DATE: 25-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K  
; REGISTRATION NUMBER: P39,300  
; REFERENCE/DOCKET NUMBER: 010830-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7566 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: plasmid DNA designated as pPS0212  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1785  
; OTHER INFORMATION: /note= "Coding region of a  
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the  
; OTHER INFORMATION: cryIab6 gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1793..2026  
; OTHER INFORMATION: /note= "3' regulatory sequence  
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu  
; OTHER INFORMATION: T-DNA gene 7."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2396..2921  
; OTHER INFORMATION: /note= "35S promoter sequence  
; OTHER INFORMATION: derived from Cauliflower mosaic virus."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2922..3581  
; OTHER INFORMATION: /note= "Coding sequence of  
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."  
; FEATURE:

; NAME/KEY: misc\_feature  
; LOCATION: 3582..4407  
; OTHER INFORMATION: /note= "3' regulatory sequence  
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacte  
; OTHER INFORMATION: T-DNA octopine synthase gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 5600..6457  
; OTHER INFORMATION: /note= "sequence complementary to  
; Patent No. 5952547  
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 7071..7566  
; OTHER INFORMATION: /note= "TR1' and TR2 promoter  
; OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader wit  
; OTHER INFORMATION: respect to sequence of pUD884 of SEQ ID NO. 22."  
US-08-232-016-23  
  
Query Match 96.0%; Score 24; DB 2; Length 7566;  
Best Local Similarity 100.0%; Pred. No. 0.0045;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 gatcccccgatgagctaagctagc 25  
|||||  
Db 1808 GATCCCCGATCAGCTAAGCTAGC 1831  
  
RESULT 13  
US-08-232-016-22  
; Sequence 22, Application US/08232016  
; Patent No. 5952547  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; APPLICANT: VAN AARSEN, Roel  
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
; TITLE OF INVENTION: PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.016  
; FILING DATE: 03-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 91402920.2  
; FILING DATE: 30-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 92400820.4  
; FILING DATE: 25-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McGowan, Malcolm K  
; REGISTRATION NUMBER: P39,300  
; REFERENCE/DOCKET NUMBER: 010830-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7566 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: plasmid DNA designated as pPS0212  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1785  
; OTHER INFORMATION: /note= "Coding region of a  
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the  
; OTHER INFORMATION: cryIab6 gene."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1793..2026  
; OTHER INFORMATION: /note= "3' regulatory sequence  
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu  
; OTHER INFORMATION: T-DNA gene 7."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2396..2921  
; OTHER INFORMATION: /note= "35S promoter sequence  
; OTHER INFORMATION: derived from Cauliflower mosaic virus."  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2922..3581  
; OTHER INFORMATION: /note= "Coding sequence of  
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."  
; FEATURE:

```

; LENGTH: 7639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORGANISM: plasmid DNA designated as pJD884
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1869
;   OTHER INFORMATION: /note= "Coding region of a
;   OTHER INFORMATION: truncated bc2 (cryIab) gene, also designated as the bc884 gen
;   NAME/KEY: misc_feature
;   LOCATION: 1877..2110
;   OTHER INFORMATION: /note= "3' regulatory sequence
;   OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
;   OTHER INFORMATION: T-DNA gene 7."
;   NAME/KEY: misc_feature
;   LOCATION: 2480..3005
;   OTHER INFORMATION: /note= "35S promoter sequence
;   OTHER INFORMATION: derived from Cauliflower mosaic virus."
;   NAME/KEY: misc_feature
;   LOCATION: 3006..3665
;   OTHER INFORMATION: /note= "Coding sequence of
;   OTHER INFORMATION: chloramphenicol acetyl transferase gene."
;   NAME/KEY: misc_feature
;   LOCATION: 3666..4491
;   OTHER INFORMATION: /note= "3' regulatory sequence
;   OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
;   OTHER INFORMATION: T-DNA octopine synthase gene."
;   NAME/KEY: misc_feature
;   LOCATION: 5684..6541
;   OTHER INFORMATION: /note= "Sequence complementary to
;   OTHER INFORMATION: Patent No. 5952547
;   OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
;   NAME/KEY: misc_feature
;   LOCATION: 7155..7639
;   OTHER INFORMATION: /note= "Trl' and TR2' promoter
;   OTHER INFORMATION: derived from Agrobacterium T-DNA."
US-08-232-016-22

```

```

Query Match          96.0%; Score 24; DB 2; Length 7639;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 gatccccgatgagctaagctagc 25
Db 1892 GATCCCCGATGAGCTAAGCTAGC 1915

```

```

RESULT 14
US-08-549-680A-5/c
; Sequence 5, Application US/08549680A
; Patent No. 5962768
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, MARCUS
; APPLICANT: REYNAERTS, ARLETTE
; APPLICANT: GOSSELE, VERONIQUE
; APPLICANT: VAN AARSEN, ROEL
; TITLE OF INVENTION: MARKER GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia

```

```

; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,680A
; FILING DATE: 16 JANUARY 1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
;   NAME: SVENSSON, LEONARD R.
;   REGISTRATION NUMBER: 30,330
;   REFERENCE/DOCKET NUMBER: 2121-0111P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 205-8000
;   TELEFAX: (703) 205-8050
;   TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 7811 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: circular
;   MOLECULE TYPE: DNA (synthetic)
;   FEATURE:
;     NAME/KEY: misc_recomb
;     LOCATION: 1..7811
;     OTHER INFORMATION: /label= vector pTRVA3
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 194..218
;     OTHER INFORMATION: /note= "T-DNA right border"
;   NAME/KEY: misc_feature
;   LOCATION: 484..684
;   OTHER INFORMATION: /note= "the 3' end formation and
;   OTHER INFORMATION: polyadenylation region of T-DNA gene 7"
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: complement (729..1340)
;     OTHER INFORMATION: /note= "the aac(6') coding
;     OTHER INFORMATION: sequence"
;     FEATURE:
;       NAME/KEY: promoter
;       LOCATION: 1341..1756
;       OTHER INFORMATION: /label= 35S promoter
;     FEATURE:
;       NAME/KEY: misc_feature
;       LOCATION: 3001..3023
;       OTHER INFORMATION: /note= "T-DNA left border
;       OTHER INFORMATION: sequences"
US-08-549-680A-5

```

```

Query Match          96.0%; Score 24; DB 2; Length 7811;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 gatccccgatgagctaagctagc 25
Db 702 GATCCCCGATGAGCTAAGCTAGC 679

```

```

RESULT 15
US-08-232-016-11/c
; Sequence 11, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke

```

APPLICANT: DOCKX, Jan  
APPLICANT: VAN AARSEN, Roel  
TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
TITLE OF INVENTION: PLANT CELLS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESS: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.016  
FILING DATE: 03-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91402920.2  
FILING DATE: 30-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92400820.4  
FILING DATE: 25-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K  
REGISTRATION NUMBER: P39,300  
REFERENCE/DOCKET NUMBER: 010830-049  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: oligonucleotide, designated as PS59  
US-08-232-016-11

Query Match 68.0%; Score 17; DB 2; Length 29;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 gatcccccgatgagcta 18  
|||||  
Db 17 GATCCCCCGATGAGCTA 1

Search completed: February 25, 2002, 18:05:15  
Job time: 18563 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:20:57 ; Search time 8261.74 Seconds  
(without alignments)  
32.517 Million cell updates/sec

Title: US-09-698-903B-5  
Perfect score: 25  
Sequence: 1 ggatcccccagatgagctaagctagc 25

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_estfun: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estom: \*  
5: em\_estpl: \*  
6: em\_estba: \*  
7: em\_estro: \*  
8: em\_estov: \*  
9: em\_htc: \*  
10: gb\_est1: \*  
11: gb\_est2: \*  
12: gb\_htc: \*  
13: gb\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_hum: \*  
16: em\_gss\_inv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_rod: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.2	72.8	498	13 AQ791259	AQ791259 HS_5498_A
2	18.2	72.8	562	13 AZ028032	AZ028032 RPCI-23-3
3	17.6	70.4	303	10 AW862033	AW862033 RC3-CT034
4	17.6	70.4	312	10 BB453584	BB453584 BB453584
5	17.6	70.4	454	11 N24161	N24161 yx96d12.s1
6	17.6	70.4	497	11 BF496895	BF496895 AT10901.5
7	17.6	70.4	510	11 BI038578	BI038578 MR4-NF014
8	17.6	70.4	587	11 BF877912	BF877912 MR0-ET010
9	17.6	70.4	605	10 BE639934	BE639934 946044B07
10	17.6	70.4	611	13 AQ834638	AQ834638 HS_5442_A
11	17.6	70.4	924	13 CNS038XY	AL23215 Tetraodon
12	17.2	68.8	295	10 BE120999	BE120999 UI-R-CA0-

13	17.2	68.8	326	10 BB499943	BB499943 BB499943
14	17.2	68.8	336	11 BF410724	UI-R-CA0-
15	17.2	68.8	368	13 AQ867402	AQ867402 nbeb0031F
16	17.2	68.8	387	10 AW501051	UI-HF-BP0
17	17.2	68.8	398	10 AI183978	AI183978 qd69g01.x
18	17.2	68.8	407	10 AA429095	AA429095 zv49d03.f
19	17.2	68.8	410	10 AW581963	AW581963 MR4-ST012
20	17.2	68.8	462	11 BI134877	UI-M-BH03-
21	17.2	68.8	501	11 BG073283	H3120G12-
22	17.2	68.8	549	13 TA311E120	AL489968 T. brucei
23	17.2	68.8	554	11 BI133671	UI-M-BH3-
24	17.2	68.8	554	13 AQ509694	AQ509694 nbdx0096F
25	17.2	68.8	568	10 BE048724	BE048724 hr52h07.x
26	17.2	68.8	574	10 BE026535	BE026535 db27g06.x
27	17.2	68.8	583	11 BI134114	UI-M-BH3-
28	17.2	68.8	585	10 BE326762	BE326762 hr64e06.x
29	17.2	68.8	622	10 AW548994	AW548994 L0045H03-
30	17.2	68.8	627	10 BE535820	BE535820 601060510
31	17.2	68.8	630	11 BG470582	BG470582 602511491
32	17.2	68.8	707	11 BG334339	BG334339 602461858
33	17.2	68.8	729	10 BE394825	BE394825 601312066
34	17.2	68.8	755	10 BE543444	BE543444 601070751
35	17.2	68.8	768	11 BG678254	BG678254 602624567
36	17.2	68.8	781	10 AI344042	AI344042 tco1905.x
37	17.2	68.8	812	11 BG035724	BG035724 602325804
38	17.2	68.8	825	11 BG166218	BG166218 602340976
39	17.2	68.8	894	11 BF301141	BF301141 602029166
40	17.2	68.8	916	11 BF797773	BF797773 602257530
41	17.2	68.8	992	13 CNS0333U	AL225651 Tetraodon
42	17.2	68.8	1498	11 BG476758	BG476758 602524680
43	17.2	68.8	1649	12 AK016762	AK016762 Mus muscu
44	17	68.0	251	10 BE419550	BE419550 WWS014.E1
45	17	68.0	400	10 BE119925	BE119925 UI-R-CA0-

#### ALIGNMENTS

RESULT	1
AQ791259	AQ791259
LOCUS	HS_5498_A2_B10_37A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION	genomic clone Plate=1074 Col=20 Row=C, DNA sequence.
ACCESSION	AQ791259
VERSION	AQ791259.1 GI:5698806
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 498) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pletier de Jong (pletier@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.hscsc.washington.edu Plate: 1074 row: C column: 20 Seq primer: T7

Class: BAC ends  
High quality sequence stop: 498.  
Location/Qualifiers  
1. 498  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Plate-1074 Col=20 Row=C"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACE3.6 vector at EcoRI sites"  
BASE COUNT 166 a 90 c 85 g 151 t 6 others  
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 498;  
Best Local Similarity 87.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcccccgatgagctaaagctagc 25  
||||| ||||| ||||| |||||  
Db 414 ATCCCCGATGCTCTAAGATAGC 436

RESULT 2  
AZ028032/c  
LOCUS  
DEFINITION  
RPCI-23-373M15-TV RPCI-23 Mus musculus genomic clone RPCI-23-373M15  
, DNA sequence.  
ACCESSION  
AZ028032  
VERSION  
AZ028032.1 GI:7103416  
KEYWORDS  
GSS.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 562)  
Zhaoh.S., Nierman.W., Feldblyum.T., Malek.J., Shatsman.S., Akinret  
B., Levins.M., Megann.S., Isegaye.G., Geer.K., Krol.M., de Jong.P.  
and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other\_GSSs: RPCI-23-373M15.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
Plate: 373 row: M column: 15  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers  
1. .562  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="RPCI-23-373M15"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 122 a 122 c 163 g 153 t 2 others  
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 562;  
Best Local Similarity 87.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcccccgatgagctaaagctagc 25  
||||| ||||| ||||| |||||  
Db 247 ATCCACCAATGAGCTTAGCTAGC 225

RESULT 3  
AW862033/c  
LOCUS  
DEFINITION  
RC3-CT0347-110300-014-a10 CT0347 Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
AW862033  
VERSION  
AW862033.1 GI:7957726  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 303)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=4t2-RC3-CT0347-110  
300-014-a10&t3=2000-03-11&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 23  
High quality sequence stop: 303.  
Location/Qualifiers  
1. 303  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0347"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 104 a 60 c 62 g 77 t  
ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 303;





Query Match	70.4%	Score 17.6;	DB 11;	Length 497;
Best Local Similarity	83.3%	Pred. No. 5.4e+02;		

RESULT	8				
BF877912		BF877912	587 bp	mRNA	EST
					17-JAN-2001

```

DEFINITION MRO-ET0109-171100-001-a01 ET0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF877912
VERSION BF877912.1 GI:12268042
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 587)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202863
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-ET0109-
171100-001-a01&t3=2000-11-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 521.
FEATURES             Location/Qualifiers
     source           1..587
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ET0109"
                     /dev_stage="Adult"
                     /note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT          149 a 131 c 164 g 143 t
ORIGIN
1 ggatcccccgcgatgaactagctag 24
|||||
15 GGATCCCCCGATGACTAATCTGTG 38

Query Match          70.4%; Score 17.6; DB 11; Length 587;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ggatcccccgcgatgaactagctag 24
|||||
15 GGATCCCCCGATGACTAATCTGTG 38

Db 15 GGATCCCCCGATGACTAATCTGTG 38

RESULT 9
LOCUS BE639934
DEFINITION 94604B07.y1 946 - tassal primordium prepared by Schmidt lab Zea
mays cDNA, mRNA sequence.
ACCESSION BE639934
VERSION BE639934.1 GI:9953351
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.

Query Match          70.4%; Score 17.6; DB 11; Length 587;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ggatcccccgcgatgaactagctag 24
|||||
15 GGATCCCCCGATGACTAATCTGTG 38

Db 15 GGATCCCCCGATGACTAATCTGTG 38

RESULT 9
LOCUS BE639934
DEFINITION 94604B07.y1 946 - tassal primordium prepared by Schmidt lab Zea
mays cDNA, mRNA sequence.
ACCESSION BE639934
VERSION BE639934.1 GI:9953351
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 605)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946044 row: B column: 07.
FEATURES             Location/Qualifiers
     source           1..605
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="946 - tassal primordium prepared by Schmidt
                     lab"
                     /tissue_type="tassels"
                     /dev_stage="just after the transition from vegetative to
                     inflorescence development"
                     /lab_host="XlOIR"
                     /note="Organ: tassels; Vector: HybrizAP; Site_1: EcoRI;
                     Site_2: XhoI; George Chuck dissected immature tassels
                     between 1mm and 3mm. Sharon Stanfield prepared the cDNA
                     library in HybrizAP. Sample insert size range was 350 bp
                     to 3 Kb with a 1 Kb average."
BASE COUNT          138 a 160 c 186 g 120 t
ORIGIN
1 ggatcccccgcgatgaactagctagc 25
|||||
533 GAATCCCCCGTAGCTAGGCAAGC 557

Query Match          70.4%; Score 17.6; DB 10; Length 605;
Best Local Similarity 80.0%; Pred. No. 5.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 ggatcccccgcgatgaactagctagc 25
|||||
533 GAATCCCCCGTAGCTAGGCAAGC 557

Db 533 GAATCCCCCGTAGCTAGGCAAGC 557

RESULT 10
LOCUS AQ834638
DEFINITION HS_5442_AL_H01_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1018 Col=1 Row=0, DNA sequence.
ACCESSION AQ834638
VERSION AQ834638.1 GI:5800700
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 611)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999).
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

```

```

FEATURES
  source
    1..924
      Location/Qualifiers
        ntlp://www.genoscope.cns.fr/Tetraodon.
        /organism="Tetraodon nigroviridis"
        /db_xref="taxon:99883"
        /clone="00608"

```

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA0-baf-h-12-0-UI"
/clone.lib="UI-R-CA0"
/lab.host="DH10B (Life Technologies)"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
, midbrain, cerebral cortex, corpus striatum, testis, and
hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at rateat.eng.uiowa.edu. The subtraction has been
previously described in (Ronaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-CA0
TAG_TISSUE=hypothalamus
TAG_SEQ=GATGC"

```

```

BASE COUNT      75 a      67 c      72 g      81 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 10; Length 295;
Best Local Similarity 86.4%; Pred. No. 7.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggatcccgatgagctaaagt 22
||||| ||| ||| |||
Db 237 GGATCCCGACGACCGAGCT 216

RESULT 13
LOCUS BB499943 326 bp mRNA EST 26-JUL-2000
DEFINITION BB499943 RIKEN full-length enriched, 0 day neonate kidney Mus
musculus cDNA clone D630023P14 3', mRNA sequence.
ACCESSION BB499943
VERSION BB499943.1 GI:9476600
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T., Carninci
P., Endo.T., Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N.,
Hirozane.T., Horii.F., Ishii.Y., Ishikawa.J., Ishikawa.T., Itoh.M.,
Izawa.M., Kadota.K., Kagawa.I., Kai.C., Kawai.J., Kikuchi.N.,
Kiyosawa.H., Kojima.Y., Kondo.S., Koya.S., Kurihara.C., Kusakabe.M.,
Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H., Okazaki.Y.,
Ono.T., Owa.C., Saito.H., Sakai.C., Sato.K., Shibata.K., Shibata
Y., Shigemoto.Y., Shinagawa.A., Shiraki.T., Sogabe.Y., Sugahara.Y.,
Suzuki.H., Suzuki.H., Tagawa.A., Takahashi.F., Tomimaga.N., Toya
T., Tsunoda.Y., Watahiki.A., Watanabe.S., Yamamura.T., Yamanaka.I.,
Yano.R., Yasunishi.A., Yokota.T., Yoshida.K., Yoshiki.A., Yoshino
M., Muramatsu.M. and Hayashizaki.Y.
RIKEN Mouse ESTs (Konno.H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuunai,T., Akiyama,J., Shibata,K., Izawa.M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu.M., Okazaki
Y. and Hayashizaki.Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..326
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D630023P14"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"

/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCGAGTTAAATTAATTCCTCCCTCCCTCCCTCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      95 a      68 c      60 g      103 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 10; Length 326;
Best Local Similarity 86.4%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tccccgatgagctaaagtacg 25
||||| ||| ||| ||| |||
Db 107 TCCCCGATGAGCGAGCTGCG 128

RESULT 14
LOCUS BF410724 336 bp mRNA EST 28-NOV-2000
DEFINITION UI-R-CA0-bmc-b-04-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
ACCESSION BF410724
VERSION BF410724.1 GI:11398699
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 336)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized thalamus library cDNA library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
source
Location/Qualifiers
1..336
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA0-bmc-b-04-0-UI"
/clone_lib="UI-R-CA0"
/lab_host="DH10B (Life Technologies)"

```

/notes=Vector: pTY73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not 1; Site\_2: Eco RI; The UI-R-CA0  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
 , midbrain, cerebral cortex, corpus striatum, testis, and  
 hippocampus. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at [rateat.eng.utoraw.edu](http://rateat.eng.utoraw.edu). The subtraction has been  
 previously described in (Bonaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996).  
 TAG\_LIB=UI-R-CA0  
 TAG\_TISSUE=thalamus  
 TAG\_SEQ=CAWCG\*

BASE COUNT	82 a	73 c	85 g	96 t
ORIGIN				

```
Query Match      68.8%; Score 17.2; DB 11; Length 336;
Best Local Similarity 86.4%; Pred. No. 7.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0
```

Qy 1 ggatccccgatgagctaagct 22  
 |||||  
 Db 238 GGATCCCCGACCGGAAGCT 217

[illegible]

ACCESSION  
AQ867402  
VERSION  
AQ867402.1 GI:6217859

SOURCE	ORGANISM	Oryza sativa.	Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

1. (bases 1 to 368)  
Wing, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293

Email: [rwing@clmson.edu](mailto:rwing@clmson.edu)  
 Seq primer: TAATACGACTCACTATAGGG  
 Class: BAC ends  
 High quality sequence start: 23  
 High quality sequence stop: 328.  
 Location/Qualifiers  
     1..368  
FEATURES  
source

source

```

/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="hebb0031f05"
/clone_lib="CGUI Rice BAC Library (EGORI)"
/tissue_type="leaf"
/lab_host="F.coli.nippon"

```

`/note=Vector: pBCIndigo; Site_1: ECORI; site_2: ECORI;`  
Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable

for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation of a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/SSP Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu))."

	BASE COUNT	ORIGIN
1	0.0000	0.0000
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0000	0.0000
5	0.0000	0.0000
6	0.0000	0.0000
7	0.0000	0.0000
8	0.0000	0.0000
9	0.0000	0.0000
10	0.0000	0.0000
11	0.0000	0.0000
12	0.0000	0.0000
13	0.0000	0.0000
14	0.0000	0.0000
15	0.0000	0.0000
16	0.0000	0.0000
17	0.0000	0.0000
18	0.0000	0.0000
19	0.0000	0.0000
20	0.0000	0.0000
21	0.0000	0.0000
22	0.0000	0.0000
23	0.0000	0.0000
24	0.0000	0.0000
25	0.0000	0.0000
26	0.0000	0.0000
27	0.0000	0.0000
28	0.0000	0.0000
29	0.0000	0.0000
30	0.0000	0.0000
31	0.0000	0.0000
32	0.0000	0.0000
33	0.0000	0.0000
34	0.0000	0.0000
35	0.0000	0.0000
36	0.0000	0.0000
37	0.0000	0.0000
38	0.0000	0.0000
39	0.0000	0.0000
40	0.0000	0.0000
41	0.0000	0.0000
42	0.0000	0.0000
43	0.0000	0.0000
44	0.0000	0.0000
45	0.0000	0.0000
46	0.0000	0.0000
47	0.0000	0.0000
48	0.0000	0.0000
49	0.0000	0.0000
50	0.0000	0.0000
51	0.0000	0.0000
52	0.0000	0.0000
53	0.0000	0.0000
54	0.0000	0.0000
55	0.0000	0.0000
56	0.0000	0.0000
57	0.0000	0.0000
58	0.0000	0.0000
59	0.0000	0.0000
60	0.0000	0.0000
61	0.0000	0.0000
62	0.0000	0.0000
63	0.0000	0.0000
64	0.0000	0.0000
65	0.0000	0.0000
66	0.0000	0.0000
67	0.0000	0.0000
68	0.0000	0.0000
69	0.0000	0.0000
70	0.0000	0.0000
71	0.0000	0.0000
72	0.0000	0.0000
73	0.0000	0.0000
74	0.0000	0.0000
75	0.0000	0.0000
76	0.0000	0.0000
77	0.0000	0.0000
78	0.0000	0.0000
79	0.0000	0.0000
80	0.0000	0.0000
81	0.0000	0.0000
82	0.0000	0.0000
83	0.0000	0.0000
84	0.0000	0.0000
85	0.0000	0.0000
86	0.0000	0.0000
87	0.0000	0.0000
88	0.0000	0.0000
89	0.0000	0.0000
90	0.0000	0.0000
91	0.0000	0.0000
92	0.0000	0.0000
93	0.0000	0.0000
94	0.0000	0.0000
95	0.0000	0.0000
96	0.0000	0.0000
97	0.0000	0.0000
98	0.0000	0.0000
99	0.0000	0.0000
100	0.0000	0.0000

Query Match 68.8%; Score 17.2; DB 13;  
Best Local Similarity 86.4%; Pred. No. 7.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0

Oy 3 atccccgatgagctaagctag 24  
||| ||| ||| ||| ||| |||  
Db 34 ATCACCCCAATGAGCTCAGCTAG 55

Search completed: February 25, 2002, 17:21:01  
Job time: 16154 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:10 ; Search time 2331.3 Seconds  
(without alignments)  
155.680 Million cell updates/sec

Title: US-09-698-903B-6

Perfect score: 22

Sequence: 1 tcctctacggcaatgtaccagc 22

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_ov.\*

21: em\_or.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	100.0	22	6	AX127753	AX127753 Sequence
2	22	100.0	22	6	AX172462	AX172462 Sequence
3	22	100.0	249	12	ARGMTUB	X05579 Soybean bet
4	22	100.0	831	1	ATRN7	V00090 Agrobacteri
5	22	100.0	878	1	ATTDNA	X00431 Agrobacteri
6	22	100.0	1037	6	A10942	A10942 Nucleotide
7	22	100.0	1085	6	A10939	A10939 Nucleotide
8	22	100.0	1160	6	A10943	A10943 Nucleotide
9	22	100.0	1166	6	A10941	A10941 Nucleotide
10	22	100.0	1186	6	A18051	A18051 DNA used as
11	22	100.0	1186	6	AR095107	AR095107 Sequence
12	22	100.0	1186	6	AR098313	AR098313 Sequence
13	22	100.0	1186	6	AX012338	AX012338 Sequence
14	22	100.0	1186	6	I49886	I49886 Sequence 2
15	22	100.0	1186	6	I82374	I82374 Sequence 2
c 16	22	100.0	2476	12	TBI251013	AJ251013 Transform
17	22	100.0	3200	6	I44104	I44104 Sequence 23
18	22	100.0	3201	6	I44103	I44103 Sequence 22
c 19	22	100.0	3236	12	TBI251014	AJ251014 Transform
c 20	22	100.0	4832	6	AX172441	AX172441 Sequence
c 21	22	100.0	4946	6	A60108	A60108 Sequence 1
c 22	22	100.0	4946	6	A76915	A76915 Sequence 1
c 23	22	100.0	4946	6	AR098307	AR098307 Sequence
c 24	22	100.0	4946	6	AX172440	AX172440 Sequence
c 25	22	100.0	5349	6	A71437	A71437 Sequence 7
c 26	22	100.0	5560	6	A60112	A60112 Sequence 5
c 27	22	100.0	5560	6	AR098311	AR098311 Sequence
28	22	100.0	5865	6	AX127748	AX127748 Sequence
c 29	22	100.0	5865	6	AX127748	AX127748 Sequence
30	22	100.0	6539	6	E31991	E31991 Mutated bar
31	22	100.0	6548	6	A60109	A60109 Sequence 2
32	22	100.0	6548	6	A76916	A76916 Sequence 2
33	22	100.0	6548	6	AR098308	AR098308 Sequence
34	22	100.0	6548	6	E31990	E31990 Mutated bar
35	22	100.0	7566	6	A24783	A24783 plasmid pPS
36	22	100.0	7566	6	AR074388	AR074388 Sequence
37	22	100.0	7599	6	AX063413	AX063413 Sequence
38	22	100.0	7639	6	A24782	A24782 plasmid pJD
39	22	100.0	7639	6	AR074387	AR074387 Sequence
c 40	22	100.0	7811	6	AR078675	AR078675 Sequence
41	22	100.0	12095	12	BINHYGDNA	237515 Binary vect
c 42	22	100.0	24595	1	ATACH5	X00493 Agrobacteri
c 43	22	100.0	24595	6	E00404	E00404 Ti plasmid
c 44	22	100.0	24595	6	E00546	E00546 DNA fragmen
c 45	22	100.0	194140	1	AF242881	AF242881 Agrobacte

ALIGNMENTS

RESULT 1						
AX127753	AX127753	22 bp	DNA	PAT	15-MAY-2001	
LOCUS	Sequence 6 from Patent WO0131042.					
DEFINITION	AX127753					
ACCESSION	AX127753.1					
VERSION	GI:14134400					
KEYWORDS	synthetic construct.					
SOURCE	synthetic construct					
ORGANISM	artificial sequence.					

1 (bases 1 to 22)

Weston,B. and de Beuckeleer,M.

Male-sterile brassica plants and methods for producing same

Patent: WO 0131042-A 6 03-MAY-2001;

Aventis CropScience N.V. (BE)

Location/Qualifiers

1. .22

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="primer MDB193"

```

BASE COUNT      6 a      7 c      4 g      5 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
    |||||
Db 1 TCATCTACGGCAATGTACCAGC 22

RESULT 2
LOCUS AX172462 22 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 23 from Patent WO0141558.
ACCESSION AX172462
VERSION AX172462.1 GI:14597574
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 22)
AUTHORS de Both, G. and de Beuckeleer, M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 23 14-JUN-2001;
Aventis CropScience N.V. (BE)
FEATURES
    source
        1..22
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="primer 193"
BASE COUNT      6 a      7 c      4 g      5 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
    |||||
Db 1 TCATCTACGGCAATGTACCAGC 22

RESULT 3
ARGMTUB
LOCUS Soybean beta-1-tubulin gene fused to Ti plasmid unit 7 3078.
DEFINITION X05579
ACCESSION X05579.1 GI:58087
VERSION beta-tubulin; fusion gene; plasmid.
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 249)
AUTHORS Guiltinan, M.J., Veiten, J., Bustos, M.M., Cyr, R.J., Schell, J. and Fosket, D.E.
TITLE The expression of a chimeric soybean beta-tubulin gene in tobacco
JOURNAL Mol. Gen. Genet. 207, 328-334 (1987)
FEATURES
    source
        1..249
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            1..51
                /note="fusion product (17AA); Protein sequence is in
                conflict with the conceptual translation"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA29084.1"
                /db_xref="GI:4376141"
                /translation="MAAAWASSNWSTDPPTS"
CDS

```

---

```

misc_feature      1..13
    /note="beta-1-tubulin sequence"
misc_feature      14..28
    /note="pUC 13 polylinker"
misc_feature      29..35
    /note="Sal I linker"
misc_feature      37
    /note="theroretical fusion junction (24) with gene 7 of T1
    plasmid"
misc_feature      173..178
    /note="put.polyA signal"
polyA_site        198
    /note="polyA site"
misc_feature      220..225
    /note="put.polyA signal"
BASE COUNT      76 a      47 c      38 g      88 t
ORIGIN

Query Match      100.0%; Score 22; DB 12; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcattcagcgcaatgtaccagc 22
    |||||
Db 106 TCATCTACGGCAATGTACCAGC 127

RESULT 4
ATTRN7
LOCUS Agrobacterium tumefaciens gene encoding 'transcript 7' mRNA for a
DEFINITION protein with unknown function.
ACCESSION V00090
VERSION V00090.1 GI:39180
KEYWORDS unidentified reading frame.
SOURCE Agrobacterium tumefaciens.
ORGANISM Agrobacterium tumefaciens.
REFERENCE 1 (bases 1 to 831)
AUTHORS Dhaese, P., De Greve, H., Gielen, J., Seurinck, J., Van Montagu, M.M.
and Schell, J.
TITLE Identification of sequences involved in the polyadenylation of
higher plant nuclear transcripts using Agrobacterium T-DNA genes as
models
JOURNAL EMBO J. 2, 419-426 (1983)
REFERENCE 2 (bases 76 to 100)
AUTHORS Dhaese, P.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1983) to the EMBL/GenBank/DBJ databases
COMMENT Data kindly reviewed (27-MAY-1983) by Dhaese P.
FEATURES
    source
        1..831
            /organism="Agrobacterium tumefaciens"
            /strain="(octopine Ti-DNA)"
            /db_xref="taxon:358"
            75..81
                /note="CG [1] revised CCAGAGG [2]"
                /citation=[1]
                /citation=[2]
                99..101
                    /note="CCA [1] revised CTA [2]"
                    /citation=[1]
                    /citation=[2]
                    132..673
                        /note="transcript 7 (alternate)"
                        132..672
                            /note="transcript 7"
                            148..528
                                /note="unknown gene (148 is 1st base in codon) (525 is 3rd
                                base in codon)"
                                /codon_start=1

```

```
/transl_table=11
/protein_id="CAA23429.1"
/db_xref="GI:39181"
/db_xref="SWISS-PROT:P03867"
/translation="MNFADTPLASLDLDWACEEFIKTYGASPOLETGEVIOTNNGLLY
LYGKLSQRHIDHILKPKKEELSFTTIKPAEMKAQQSDLTYYVAIFQSNVFLCVSN
PEKGFRCNRPFLYPPIVAGSMS"
BASE COUNT      262 a 176 c 138 g 255 t
ORIGIN

Query Match      100.0%; Score 22; DB 1; Length 831;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 580 TCATCTACGCGCAATGTACCAGC 601

RESULT 5
ATTDNA
LOCUS      878 bp      DNA      BCT      25-MAR-1996
DEFINITION Agrobacterium tumefaciens crown gall tumor T-DNA from Ti (tumor
            inducing) plasmid pTiA6.
ACCESSION X00431.1 GI:39150
VERSION   X00431.1
KEYWORDS plasmid.
SOURCE   Agrobacterium tumefaciens.
ORGANISM Agrobacterium tumefaciens.
REFERENCE 1 (bases 1 to 878)
AUTHORS   McPherson,J.C.
TITLE     DNA sequence analysis of crown gall tumor T-DNA encoding the 0.7 kb
transcript
JOURNAL   Nucleic Acids Res. 12 (5), 2317-2325 (1984)
MEDLINE   84169535
FEATURES
    source
        1..878
            Location/Qualifiers
                organism="Agrobacterium tumefaciens"
                strain="plasmid pTiA6"
                db_xref="taxon:358"
                note="TATA-box"
                note="polyadenylation signal"
                note="unidentified reading frame"
                codon_start=1
    promoter
        60..66
    misc_feature
        68..613
        note="polyadenylation signal"
    CDS
        109..489
            /note="unidentified reading frame"
            /transl_table=11
            /protein_id="CAA25129.1"
            /db_xref="GI:39151"
            /db_xref="SWISS-PROT:P03867"
            /translation="MNFADTPLASLDLDWACEEFIKTYGASPOLETGEVIOTNNGLLY
            LYGKLSQRHIDHILKPKKEELSFTTIKPAEMKAQQSDLTYYVAIFQSNVFLCVSN
            PEKGFRCNRPFLYPPIVAGSMS"
            659..664
    misc_feature
        659..664
            /note="polyadenylation signal"
BASE COUNT      288 a 189 c 139 g 262 t
ORIGIN

Query Match      100.0%; Score 22; DB 1; Length 878;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 541 TCATCTACGCGCAATGTACCAGC 562

RESULT 6
```

```
A10942
LOCUS      A10942      1037 bp      DNA      PAT      27-SEP-1993
DEFINITION Nucleotide sequence 4 from patent number DE3920034.
ACCESSION  A10942
VERSION    A10942.1 GI:492369
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 1037)
AUTHORS    Patent: DE 3920034-A 4 31-MAY-1990;
JOURNAL    Location/Qualifiers
FEATURES    1..1037
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT      338 a 174 c 166 g 359 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 1037;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 837 TCATCTACGCGCAATGTACCAGC 858

RESULT 7
A10939
LOCUS      A10939      1085 bp      DNA      PAT      27-SEP-1993
DEFINITION Nucleotide sequence 1 from patent number DE3920034.
ACCESSION  A10939
VERSION    A10939.1 GI:492367
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 1085)
AUTHORS    Patent: DE 3920034-A 1 31-MAY-1990;
JOURNAL    Location/Qualifiers
FEATURES    1..1085
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT      369 a 218 c 155 g 343 t
ORIGIN

Query Match      100.0%; Score 22; DB 6; Length 1085;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 885 TCATCTACGCGCAATGTACCAGC 906

RESULT 8
A10943
LOCUS      A10943      1160 bp      DNA      PAT      27-SEP-1993
DEFINITION Nucleotide sequence 5 from patent number DE3920034.
ACCESSION  A10943
VERSION    A10943.1 GI:492370
KEYWORDS   .
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 1160)
AUTHORS    Patent: DE 3920034-A 5 31-MAY-1990;
JOURNAL    Location/Qualifiers
FEATURES
```

```

source
1. .1160
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 367 a 194 c 188 g 411 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1160;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 960 TCATCTACGGCAATGTACCAGC 981

RESULT 9
AL0941
LOCUS AR0941 1166 bp DNA PAT 27-SEP-1993
DEFINITION Nucleotide sequence 3 from patent number DE3920034.
ACCESSION AR0941
VERSION AR0941.1 GI:492368
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1166)
AUTHORS
JOURNAL
FEATURES
source
BASE COUNT 381 a 208 c 196 g 381 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1166;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 1066 TCATCTACGGCAATGTACCAGC 1087

RESULT 10
AL8051
LOCUS AR8051 1186 bp DNA PAT 26-JUL-1994
DEFINITION DNA used as a probe for neo gene seq ID No:2.
ACCESSION AR8051
VERSION AR8051.1 GI:593120
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1186)
AUTHORS
JOURNAL
FEATURES
source
Patent: WO 9209696-A 2 11-JUN-1992;
Location/Qualifiers
1. .1186
/organism="synthetic construct"
/db_xref="taxon:32630"
1_8
/promoter
/note="sequence derived from tapetum specific promoter of
Nicotiana tabacum"
167.790
gene
/gene="neomycine phosphotransferase gene"
167.790
CDS
/gene="neomycine phosphotransferase gene"
/note="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1

/transl_table=11
/protein_id="CAA01373.1"
/db_xref="GI:4529900"
/translation="NELODEAARLSWLATTGVPCAAYLDVVTEAGRDWLLIGVEVPGQD
LSSHLPAAEKVSTIMADMRRLHTLDPATCFDQAKHRIERARTMEAGLVQDDLD
EEHQGLAPAEFLKARMPDGEDLVVTHGDACLPLNIMVENGFRSGFDGRLGVADR
YQDIALATRDIAEELGGEWADRFVLVYGAAPDSQRTAFYRLDDEF"
1055.1186
/note="3' regulatory sequence containing the
polyadenylation site derived from agrobacterium T-DNA gene
7"
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 11
AR095107
LOCUS AR095107 1186 bp DNA PAT 08-SEP-2000
DEFINITION Sequence 2 from patent US 6002070.
ACCESSION AR095107
VERSION AR095107.1 GI:10022665
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1186)
AUTHORS D'Halluin,K. and Gobel,E.
TITLE Process for transforming monocotyledonous plants
JOURNAL Patent: US 6002070-A 2 14-DEC-1999;
FEATURES
Location/Qualifiers
1. .1186
source
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 12
AR098313
LOCUS AR098313 1186 bp DNA PAT 14-FEB-2001
DEFINITION Sequence 2 from patent US 6074877.
ACCESSION AR098313
VERSION AR098313.1 GI:12807570
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1186)
AUTHORS D'Halluin,K. and Gobel,E.
TITLE Process for transforming monocotyledonous plants
JOURNAL Patent: US 6074877-A 2 13-JUN-2000;
FEATURES
Location/Qualifiers
1. .1186
source
BASE COUNT 244 a 317 c 325 g 300 t
ORIGIN

```

```
Query Match          100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 13
AX012338          1186 bp      DNA          PAT          06-SEP-2000
LOCUS             Sequence 2 from Patent EP0955371.
ACCESSION         AX012338
VERSION           AX012338.1 GI:9998387
KEYWORDS           synthetic construct.
SOURCE             synthetic construct.
ORGANISM           artificial sequence.
REFERENCE          1 (bases 1 to 1186)
AUTHORS            D'Halluin,K. and Goebel,E.D.
TITLE              Process for transforming monocotyledonous plants
JOURNAL            Patent: EP 0955371-A 2 10-NOV-1999;
PLANT GENETIC SYSTEMS NV (BE)
FEATURES            Location/Qualifiers
Source             1..1186
                   /organism="synthetic construct"
                   /db_xref="taxon:32630"
                   /note="DNA used as probe for neo gene"
misc_feature       1..8
                   /note="sequence derived from tapetum specific promoter of
                   Nicotiana tabacum"
misc_feature       9..790
                   /note="coding sequence of neomycine phosphotransferase"
misc_feature       791..1186
                   /note="3' regulatory sequence containing the
                   polyadenylation site derived from Agrobacterium T-DNA gene"
BASE COUNT        244 a 317 c 325 g 300 t
ORIGIN

Query Match          100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 14
I49886           1186 bp      DNA          PAT          07-OCT-1997
LOCUS             Sequence 2 from patent US 5641664.
ACCESSION         I49886
VERSION           I49886.1 GI:2472106
KEYWORDS           Unknown.
SOURCE             Unknown.
ORGANISM           Unclassified.
REFERENCE          1 (bases 1 to 1186)
AUTHORS            D'Halluin,K. and Goebel,E.
TITLE              Process for transforming monocotyledonous plants
JOURNAL            Patent: US 5641664-A 2 24-JUN-1997;
FEATURES            Location/Qualifiers
Source             1..1186
                   /organism="unknown"
BASE COUNT        244 a 317 c 325 g 300 t
ORIGIN

Query Match          100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

RESULT 15
I82374           1186 bp      DNA          PAT          10-JUN-1998
LOCUS             Sequence 2 from patent US 5712135.
ACCESSION         I82374
VERSION           I82374.1 GI:3210671
KEYWORDS           Unknown.
SOURCE             Unknown.
ORGANISM           Unclassified.
REFERENCE          1 (bases 1 to 1186)
AUTHORS            D'Halluin,K. and Gobel,E.
TITLE              Process for transforming monocotyledonous plants
JOURNAL            Patent: US 5712135-A 2 27-JAN-1998;
FEATURES            Location/Qualifiers
Source             1..1186
                   /organism="unknown"
BASE COUNT        244 a 317 c 325 g 300 t
ORIGIN

Query Match          100.0%; Score 22; DB 6; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050

Search completed: February 25, 2002, 18:00:10
Job time: 18418 sec
```

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:24 ; Search time 716.55 Seconds  
(without alignments)  
26.322 Million cell updates/sec

Title: US-09-698-903B-6

Perfect score: 22  
Sequence: 1 tcattctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	22	100.0	22	AAH25438	PCR primer for the
2	22	100.0	22	AAH25438	PCR primer MDB193
3	22	100.0	1037	11 AAQ04705	USP-Promoter-casse
4	22	100.0	1085	11 AAQ04703	Legumin-signalpept
5	22	100.0	1160	11 AAQ04706	USP-signalpeptide
6	22	100.0	1166	11 AAQ04704	USP-Promoter-casse
7	22	100.0	1186	13 AAQ25707	Chimeric neo gene
8	22	100.0	1303	17 AAT39337	Plasmid pTS88 (Eco
9	22	100.0	3153	21 AAZ29122	Plasmid pSV131 comp
10	22	100.0	3201	12 AAQ14529	pVE36 Bt ICP codin
11	22	100.0	3201	12 AAQ15144	

12	22	100.0	3336	21	AAZ29121	Plasmid pSV130 comp
13	22	100.0	3694	21	AAZ29124	Plasmid pSV133 used
14	22	100.0	3877	21	AAZ29123	Plasmid pSV132 used
15	22	100.0	4832	22	AAH25423	Nucleotide sequenc
16	22	100.0	4946	18	AAT59531	T-DNA of plasmid p
17	22	100.0	4946	22	AAH25422	Nucleotide sequenc
18	22	100.0	5228	22	AAH25422	Plasmid pTS172delt
19	22	100.0	5349	19	AAV23239	T-DNA of pTR524.
20	22	100.0	5864	17	AAT39339	Plasmid pTCO113 T-
21	22	100.0	5864	17	AAT39339	Plasmid pTCO113 T-
22	22	100.0	5865	22	AAH25422	Chimeric T-DNA of
23	22	100.0	5865	22	AAH25422	Chimeric T-DNA of
24	22	100.0	6539	21	AAZ291097	E. coli plasmid pT
25	22	100.0	6548	17	AAT39336	Plasmid pTS174 use
26	22	100.0	6548	18	AAT61394	Plasmid pTS172. C
27	22	100.0	6548	21	AAZ91096	E. coli plasmid pT
28	22	100.0	7492	22	AAH25438	Plasmid pTS346. U
29	22	100.0	7566	14	AAQ42160	Plasmid pPS0212 co
30	22	100.0	7599	22	AAH25438	Nucleotide sequenc
31	22	100.0	7639	14	AAQ42159	Plasmid pTD884 con
32	22	100.0	24593	6	AAH50226	Sequence of opine
33	22	100.0	24596	6	AAH50182	Complete nucleotid
34	21	95.5	21	22	AAH89355	T-DNA right border
35	18	81.8	19	20	AAZ28157	Transformed Arabid
36	18	81.8	20	22	AAH25428	PCR primer for the
37	18	81.8	20	22	AAH25428	PCR primer MDB258
38	18	81.8	24	21	AAZ45415	PCR primer T4 used
39	18	81.8	415	22	AAH25438	Right (5') border
40	18	81.8	416	22	AAH25438	Left (3') border f
41	17.4	79.1	1077	22	AAH25438	Right flanking reg
42	17.2	78.2	1565	21	AAH25438	Fusarium venenatum
43	16.4	74.5	2558	13	AAQ22726	Human liver GPI-PL
44	16.4	74.5	2558	16	AAQ90583	Glycosyl-phosphati
45	16.4	74.5	14244	20	AAH25051	Polynucleotide seq

ALIGNMENTS

RESULT 1

AAH25438	
ID	AAH25438 standard; DNA; 22 BP.
XX	
AC	AAH25438;
XX	
DT	22-AUG-2001 (first entry)
XX	
DE	PCR primer for the right flanking region in transgenic plant BN-RF1.
XX	
KW	Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
KW	fertility restorer gene; barnase gene; barstar gene; PCR primer; ss.
XX	
OS	Synthetic.
XX	
PN	WO200141558-A1.
XX	
PD	14-JUN-2001.
XX	
PF	06-DEC-2000; 2000WO-EF12872.
XX	
PR	08-DEC-1999; 99US-0457037.
XX	
PA	(AVET ) AVENTIS CROPS SCIENCE NV.
XX	
PI	De Both G, De Beuckeleer M;
XX	
DR	WPI; 2001-381419/40.
XX	
PT	Transgenic winter oilseed rape plants suited for producing hybrid seed
PT	with improved qualities, comprises a male-sterility gene and fertility
XX	restorer gene, integrated into the genome -
PS	Claim 60; Page 46; 98pp; English.

CC The specification describes a pair of transgenic winter oilseed rape  
 CC plants suited for producing hybrid seed. One of the plants has an  
 CC expression cassette comprising a male-sterility gene (e.g. barnase  
 CC gene), and the other plant has an expression cassette comprising a  
 CC fertility restorer gene (e.g. barstar gene), integrated into the genome.  
 CC The fertility restorer gene is capable of preventing the activity of the  
 CC male-sterility gene. The plant pair is useful for producing hybrid seed.  
 CC Plants developed from the hybrid seed have agronomic performance,  
 CC genetic stability and adaptability to different genetic backgrounds.  
 CC The present PCR primer was used to amplify the right flanking region of  
 CC a vector in a transgenic plant which carries the TA29-barstar  
 CC transgene.

SQ Sequence 22 BP; 6 A; 7 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.034;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22  
 |||||  
 Db 1 tcattcagcgcaatgtaccagc 22

## RESULT 2

AA006995  
 ID AAD06995 standard; DNA; 22 BP.

AC AAD06995;

DT 06-AUG-2001 (first entry)

DE PCR primer MDS193 to generate the flanking region of elite event MS-B2.

KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
 KW male-sterility gene; PCR primer; thermal asymmetric interlaced;  
 KW TAIL; ss.

OS Agrobacterium sp.

PN WO200131042-A2.

PD 03-MAY-2001.

PF 26-OCT-2000; 2000WO-EP10680.

PR 29-OCT-1999; 99US-0430497.

PA (AVET ) AVENTIS CROPS SCIENCE NV.

PI Weston B, De Beuckeleer M;

DR WPI; 2001-300517/31.

PT Transgenic Brassica plants, seeds, cells or tissues, characterized by  
 PT harboring specific transformation events, particularly by presence of  
 PT male-sterility gene, at specific location in its genome -

PS Example 3; Page 28; 53pp; English.

XX The present invention relates to a transgenic Brassica plant or its  
 CC seed, cells or tissues, characterised by harbouring a specific  
 CC transformation event, particularly by the presence of a male-sterility  
 CC gene, at a specific location in the Brassica genome. Transgenic  
 CC Brassica plant is useful for producing a hybrid seed by crossing the  
 CC transgenic plant with a male-fertile Brassica plant and harvesting the  
 CC hybrid seed from the transgenic Brassica plant.  
 CC The present sequence is secondary thermal interlaced (TAIL)-PCR primer  
 CC MDS193 used to right (5') border flanking region of elite event MS-B2.  
 CC This primer corresponds to position 226-247 of plasmid pCOL13.

SQ Sequence 22 BP; 6 A; 7 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 22; DB 22; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.034;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22  
 |||||  
 Db 1 tcattcagcgcaatgtaccagc 22

## RESULT 3

AA004705  
 ID AA004705 standard; DNA; 1037 BP.

AC AA004705;

DT 12-OCT-1990 (first entry)

DE USP-Promoter-cassette USP-Pr.T7.1.

KW Foreign DNA incorporation; recombinant DNA techniques;  
 KW higher plant genome; legumin; USP-Pr.T7-1; ss.

PN DE3920034-A.

PD 31-MAY-1990.

PF 20-JUN-1989; 89DE-3920034.

PR 19-SEP-1988; 88DD-0319887.

XX (PFLA-) VE KOMB PFLANZENZUC.

PI Bassuner R, Baumlein H, Muntz K, Hai NV, Wobus U;

DR WPI; 1990-172459/23.

PT Incorporation of DNA into higher plant genome - by specified  
 PT recombinant DNA techniques.

PS Disclosure; ; pp; German.

CC The unique BglII-Ort (720-725) site is for ligating foreign DNA and the  
 CC HindIII-Ort in the 3' polylinker (1032-1037) for cloning the  
 CC cassette in the T1-vector pGA471. Agrobacterium tumefaciens is  
 CC transfected.

CC See also AA004703-Q04706.

XX SQ Sequence 1037 BP; 338 A; 174 C; 166 G; 359 T; 0 other;

Query Match 100.0%; Score 22; DB 11; Length 1037;  
 Best Local Similarity 100.0%; Pred. No. 0.059;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22  
 |||||  
 Db 837 tcattcagcgcaatgtaccagc 858

## RESULT 4

AA004703  
 ID AA004703 standard; DNA; 1085 BP.

AC AA004703;

DT 12-OCT-1990 (first entry)

DE Legumin-signalpeptide cassette Le-Sig.T7.

XX Foreign DNA incorporation; recombinant DNA techniques;





Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcactcagcgcaatgtaccagc 22  
|||||  
Db 1066 tcactcagcgcaatgtaccagc 1087

RESULT 7  
AAQ25707  
ID AAQ25707 standard; DNA; 1186 BP.  
XX  
AC AAQ25707;  
XX  
DT 07-DEC-1992 (first entry)  
XX  
DE Chimeric neo gene probe.  
XX  
KW Nicotiana tabacum; neomycine phosphotransferase; Agrobacterium; ss.  
XX  
OS Synthetic.

FH Key Location/Qualifiers  
FT promoter 1..8  
FT /tag= a  
FT /note= "sequence derived from tapetum specific  
FT promoter of Nicotiana tabacum"

FT CDS 9..790  
FT /tag= b  
FT /product= neomycine\_phosphotransferase  
FT 791..1186  
FT /tag= c  
FT /note= "3', regulatory sequence contg. the  
FT polyadenylation site derived from  
FT Agrobacterium T-DNA gene 7"

XX  
PN W09209696-A.  
XX  
PD 11-JUN-1992.  
XX  
PF 21-NOV-1991; 91WO-EP02198.  
XX  
PR 23-NOV-1990; 90EP-0403332.  
XX 08-JUL-1991; 91EP-0401888.  
XX  
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
XX  
PI Dhalluin K, Goebel E;  
XX WPI; 1992-217075/26.  
XX  
PT Transforming monocotyledonous plants e.g. cereals - comprises  
PT wounding and/or degrading cells of intact plant tissue or  
PT embryogenic callus

XX Disclosure; Page 60; 76pp; English.  
XX  
CC Two transformed corn plants were analysed by means of Southern  
CC hybridisation. As a probe, a 1184 bp EcoRI-HindIII fragment derived  
CC from another plasmid was used. The sequence of that plasmid is  
CC given below. Results showed that at least a chimeric neo gene was  
CC integrated into the plant genomic DNA.  
XX

SQ Sequence 1186 BP; 244 A; 317 C; 325 G; 300 T; 0 other;

Query Match 100.0%; Score 22; DB 13; Length 1186;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcactcagcgcaatgtaccagc 22  
|||||  
Db 1029 tcactcagcgcaatgtaccagc 1050

RESULT 8  
AAT39337  
ID AAT39337 standard; DNA; 1303 BP.  
XX  
AC AAT39337;  
XX  
DT 22-JAN-1997 (first entry)  
XX  
DE Plasmid pTS88 (EcoRI-HindIII fragment).  
XX  
KW Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;  
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1..35  
FT /tag= a  
FT /label= pGEM2  
FT /note= "polylinker of pGEM2"  
FT promoter 36..694  
FT /tag= b  
FT /label= P35S  
FT /function= 35S promoter of cauliflower mosaic virus  
FT strain CM1841  
FT CDS 695..967  
FT /tag= c  
FT /label= barstar  
FT /product= Bacillus amyloliquefaciens barstar  
FT 968..1287  
FT /tag= d  
FT /label= 3'g7  
FT /function= region containing polyadenylation signal  
FT of gene 7 og Agrobacterium T-DNA  
FT misc\_feature 1288..1303  
FT /tag= e  
FT /label= pGEM2  
FT /note= "polylinker of pGEM2"  
XX  
PN W09626283-A1.  
XX  
PD 29-AUG-1996.  
XX  
PF 21-FEB-1996; 96WO-EP00722.  
XX  
PR 21-FEB-1995; 95EP-0400364.  
XX  
PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
XX  
PI Botterman J, Cornelissen M, Michiels F;  
XX WPI; 1996-402373/40.  
XX  
PT Prodn. of male sterile plants by transforming with a chimeric  
PT construct - comprising a male sterility DNA e.g. barnase and a  
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful  
PT for generating hybrid cultivars  
XX  
PS Example 1; Page 38; 56pp; English.  
XX  
CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains  
CC barstar DNA under control of a 35S promoter. The plasmid was  
CC used with pTS174 (see also AAT39336) contg. barnase DNA under  
CC control of the stamen-specific promoter EI to produce male sterile  
CC rice cv. Koshihikari transgenic plants, and with plasmid pV5136  
CC (see also AAT39336) contg. barnase DNA under control of the stamen-  
CC specific pCA55 promoter to produce male sterile maize plants.  
CC Expression of barnase (a ribonuclease) in the stamen leads to male  
CC sterility. Constitutive expression of barstar counteracts possible  
CC low level expression of barnase DNA in non-stamen tissue.  
XX

SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 22; DB 17; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.061;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||

Db 1094 tcattacggcaatgtaccagc 1115  
|||||

## RESULT 9

AAZ29122  
ID AAZ29122 standard; DNA; 3153 BP.

XX AC AAZ29122;

XX DT 21-FEB-2000 (first entry)

XX DE Plasmid DV131 comprising L3/EGFP:NPTII/Tr7 expression cassette.

XX KW Transgenic seed; marker: aleurone-specific promoter; Plasmid DV131;  
KW GFP:NPTII fusion protein construct; L3/EGFP:NPTII/Tr7; maize R gene;  
KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene; assay;  
KW Green fluorescent protein; GFP; gene fusion; selection; screening;  
KW expression; automated seed screening technique; screenable marker;  
KW transformant; embryogenic tissue; implementation; ds.

XX OS Synthetic.

XX PN W09960129-A1.

XX PD 25-NOV-1999.

XX PF 18-MAY-1999; 99WO-US11023.

XX PR 18-MAY-1998; 98US-0080625.

XX PA (DEKA-) DEKALB GENETICS CORP.

XX PI Kriz AL, Spencer TW;

XX DR WPI; 2000-072441/06.

XX PT Screenable marker genes useful for identification of transgenic seeds  
XX for plant breeding -

XX PS Example 1; Page 164-166; 182pp; English.

XX CC The present DNA sequence is the plasmid DV131, that is used in the  
CC generation of GFP:NPTII fusion protein constructs. It contains an  
CC expression cassette comprising, a promoter from the maize L3 oleosin  
CC gene, the coding sequence of EGFP:NPTII translational fusion, excised  
CC from DV126 and the Tr7 terminator. This plasmid is used to carry a  
CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
CC operably linked to the gene encoding a screenable marker, like Green  
CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
CC allow both selection and screening of transformants. The aleurone-  
CC specific promoters direct the expression of the marker genes in  
CC embryogenic tissues, allowing selection and screening of viable  
CC transgenic seeds. Screening of transgenic seeds avoids the need for  
CC growing and assaying of seeds for transgenes and allows implementation  
CC of automated seed screening techniques for the identification of  
XX transgenic seeds.

XX SQ Sequence 3153 BP; 801 A; 857 C; 756 G; 739 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3153;  
Best Local Similarity 100.0%; Pred. No. 0.069;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||

Db 2675 tcattacggcaatgtaccagc 2696  
|||||

## RESULT 10

AAQ14529

ID AAQ14529 standard; DNA; 3201 BP.

XX AC AAQ14529;

XX DT 27-JAN-1992 (first entry)

XX DE pPS029 Bt ICP coding sequence.

XX KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
KW deletion; ss.

XX OS Synthetic.

XX PN W09116432-A.

XX PD 31-OCT-1991.

XX PF 17-APR-1991; 91WO-EP00733.

XX PR 18-APR-1990; 90EP-0401055.

XX PA (PLAN-) PLANT GENETIC SYST.

XX PI Cornelissen M, Soetaert P, Stam M, Dockx J;

XX DR WPI; 1991-339820/46.

XX PT Modified Bacillus thuringiensis insecticidal crystal protein  
XX genes - having A and T sequences changed to G and C sequences  
XX encoding same amino acids, for increased expression levels

XX PS Disclosure; Fig 6(C); 78pp; English.

XX CC "n" in the sequence refers to not known nucleotides.

XX CC pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-  
XX terminal modification and the internal modification of the Bt ICP  
XX coding sequence.

XX CC See also AAQ14529, AAQ15142-44.

XX SQ Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;

Query Match 100.0%; Score 22; DB 12; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 0.069;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
|||||

Db 2948 tcattacggcaatgtaccagc 2969  
|||||

## RESULT 11

AAQ15144

ID AAQ15144 standard; DNA; 3201 BP.

XX AC AAQ15144;

XX DT 27-JAN-1992 (first entry)

XX DE pVE36 Bt ICP coding sequence.

XX KW Bacillus thuringiensis; insecticidal crystal protein; ICP;  
KW deletion; ss.

XX OS Synthetic.

XX PN WO9116432-A.  
 XX PD 31-OCT-1991.  
 XX PF 17-APR-1991; 91WO-EP00733.  
 XX PR 18-APR-1990; 90EP-0401055.  
 XX PA (PLAN-) PLANT GENETIC SYST.  
 XX PI Cornelissen M, Soetaert P, Stam M, Dockx J;  
 XX DR WPI; 1991-339820/46.  
 XX PT Modified *Bacillus thuringiensis* insecticidal crystal protein  
 PT genes - having A and T sequences changed to G and C sequences  
 PT encoding same amino acids, for increased expression levels  
 XX PS Disclosure; Fig 6(C); 78pp; English.  
 CC "n" in the sequence refers to not known nucleotides.  
 CC PPS029 (AAQ14529) is identical to pVE36, but carries both the amino-  
 CC terminal modification and the internal modification of the Bt ICP  
 CC coding sequence.  
 CC See also AAQ14529, AAQ15142-44.  
 XX SQ Sequence 3201 BP; 915 A; 651 C; 700 G; 930 T; 5 other;  
 Query Match 100.0%; Score 22; DB 12; Length 3201;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcattctacggcaatgtaccagc 22  
 Db 3021 tcattctacggcaatgtaccagc 3042  
 RESULT 12  
 AAZ29121  
 ID AAZ29121 standard; DNA; 3336 BP.  
 XX AC AAZ29121;  
 XX DT 21-FEB-2000 (first entry)  
 XX DE Plasmid DV130 comprising L3/MGFP:NPTII/Tr7 expression cassette.  
 XX KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV130;  
 KW GFP:NPTII fusion protein construct; L3/MGFP:NPTII/Tr7; maize R gene;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;  
 KW Green fluorescent protein; GFP; gene fusion; selection; screening;  
 KW transformant; expression; automated seed screening technique; assay;  
 KW screenable marker; embryogenic tissue; implementation; ds.  
 XX OS Synthetic.  
 XX PN WO9960129-A1.  
 XX PD 25-NOV-1999.  
 XX PF 18-MAY-1999; 99WO-US11023.  
 XX PR 18-MAY-1998; 98US-0080625.  
 XX PA (DEKA-) DEKALB GENETICS CORP.  
 XX PI Kriz AL, Spencer TM;  
 XX DR WPI; 2000-072441/06.  
 XX PT Screenable marker genes useful for identification of transgenic seeds  
 XX for plant breeding -  
 XX Example 1; Page 168-170; 182pp; English.  
 CC The present DNA sequence is the plasmid DV133, comprising

PT Screenable marker genes useful for identification of transgenic seeds  
 PT for plant breeding -  
 XX Example 1; Page 163-164; 182pp; English.  
 XX CC The present DNA sequence is the plasmid DV130, that is used in the  
 CC generation of GFP:NPTII fusion protein constructs. It contains an  
 CC expression cassette comprising, a promoter from the maize L3 oleosin  
 CC gene, the coding sequence of MGFP:NPTII translational fusion, excised  
 CC from DV127 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryonic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.  
 XX SQ Sequence 3336 BP; 888 A; 830 C; 764 G; 854 T; 0 other;  
 Query Match 100.0%; Score 22; DB 21; Length 3336;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcattctacggcaatgtaccagc 22  
 Db 2858 tcattctacggcaatgtaccagc 2879  
 RESULT 13  
 AAZ29124  
 ID AAZ29124 standard; DNA; 3694 BP.  
 XX AC AAZ29124;  
 XX DT 21-FEB-2000 (first entry)  
 XX DE Plasmid DV133 used for construction of GFP:NPTII fusion protein.  
 XX KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV133;  
 KW GFP:NPTII fusion protein construct; L3/RACII intron/EGFP:NPTII/Tr7;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW EGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;  
 KW Green fluorescent protein; GFP; gene fusion; maize R gene; selection;  
 KW screenable marker; screening; automated seed screening technique; assay;  
 KW transformant; expression; embryogenic tissue; implementation; ds.  
 XX OS Synthetic.  
 XX PN WO9960129-A1.  
 XX PD 25-NOV-1999.  
 XX PF 18-MAY-1999; 99WO-US11023.  
 XX PR 18-MAY-1998; 98US-0080625.  
 XX PA (DEKA-) DEKALB GENETICS CORP.  
 XX PI Kriz AL, Spencer TM;  
 XX DR WPI; 2000-072441/06.  
 XX PT Screenable marker genes useful for identification of transgenic seeds  
 XX for plant breeding -  
 XX Example 1; Page 168-170; 182pp; English.  
 CC The present DNA sequence is the plasmid DV133, comprising

CC L3/rACT1 Intron/EGFP:NPTII/Tr7, that is used in the generation of  
 CC GFP:NPTII fusion protein constructs. It contains an expression cassette  
 CC comprising, a promoter from the maize L3 oleosin gene, the rice actin 1  
 CC intron, the coding sequence of EGFP:NPTII translational fusion, excised  
 CC from DV126 and the Tr7 terminator. This plasmid is used to carry a  
 CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.  
 XX  
 SQ Sequence 3694 BP; 890 A; 984 C; 914 G; 906 T; 0 other;

Query Match 100.0%; Score 22; DB 21; Length 3694;  
 Best Local Similarity 100.0%; Pred. No. 0.07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcgaatgtaccagc 22  
 ||||||||||||||||||||  
 Db 3216 tcattctacggcgaatgtaccagc 3237

RESULT 14  
 AAZ29123  
 ID AAZ29123 standard; DNA; 3877 BP.  
 XX  
 AC AAZ29123;  
 XX  
 DT 21-FEB-2000 (first entry)  
 XX  
 DE Plasmid DV132 used for construction of GFP:NPTII fusion protein.  
 XX  
 KW Transgenic seed; marker; aleurone-specific promoter; Plasmid DV132;  
 KW GFP:NPTII fusion protein construct; L3/rACT1 intron/MGFP:NPTII/Tr7;  
 KW expression cassette; maize L3 oleosin gene; chimeric gene; transgene;  
 KW MGFP:NPTII translational fusion; Tr7 terminator; luciferase gene;  
 KW Green fluorescent protein; GFP; gene fusion; maize R gene; selection;  
 KW screenable marker; screening; automated seed screening technique; assay;  
 KW transformant; expression; embryogenic tissue; implementation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9960129-A1.  
 XX  
 PD 25-NOV-1999.  
 XX  
 PF 18-MAY-1999; 99WO-US11023.  
 XX  
 PR 18-MAY-1998; 98US-0080625.  
 XX  
 PA (DEKA-) DEKALB GENETICS CORP.  
 XX  
 PI Kriz AL, Spencer TM;  
 XX  
 DR WPI; 2000-072441/06.  
 XX  
 XX Screenable marker genes useful for identification of transgenic seeds  
 PT for plant breeding -  
 PT  
 XX Example 1; Page 166-168; 182pp; English.  
 PS  
 XX The present DNA sequence is the plasmid DV132, comprising  
 CC L3/rACT1 Intron/MGFP:NPTII/Tr7, that is used in the generation of  
 CC GFP:NPTII fusion protein constructs. It contains an expression cassette  
 CC comprising, a promoter from the maize L3 oleosin gene, the rice actin 1  
 CC intron, the coding sequence of MGFP:NPTII translational fusion, excised  
 CC from DV127 and the Tr7 terminator. This plasmid is used to carry a

CC chimeric gene, comprising an aleurone-specific promoter like L3, that is  
 CC operably linked to a gene encoding a screenable marker, like Green  
 CC fluorescent protein (GFP), luciferase or maize R gene. The gene fusions  
 CC allow both selection and screening of transformants. The aleurone-  
 CC specific promoters direct the expression of the marker genes in  
 CC embryogenic tissues, allowing selection and screening of viable  
 CC transgenic seeds. Screening of transgenic seeds avoids the need for  
 CC growing and assaying of seeds for transgenes and allows implementation  
 CC of automated seed screening techniques for the identification of  
 CC transgenic seeds.  
 XX  
 SQ Sequence 3877 BP; 977 A; 957 C; 922 G; 1021 T; 0 other;  
 Query Match 100.0%; Score 22; DB 21; Length 3877;  
 Best Local Similarity 100.0%; Pred. No. 0.071;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 tcattctacggcgaatgtaccagc 22  
 ||||||||||||||||||||  
 Db 3399 tcattctacggcgaatgtaccagc 3420  
 RESULT 15  
 AAH25423/C  
 ID AAH25423 standard; DNA; 4832 BP.  
 XX  
 AC AAH25423;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 DE Nucleotide sequence of plasmid pTHW118.  
 XX  
 KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;  
 KW fertility restorer gene; barstar gene; ss.  
 XX  
 OS Synthetic.  
 OS Streptomyces hygroscopicus.  
 OS Arabidopsis thaliana.  
 OS Bacillus amyloliquefaciens.  
 OS Nicotiana tabacum.  
 XX  
 FT Key Location/Qualifiers  
 FT misc\_feature 1..25  
 FT /tag= a  
 FT /note= "Right border repeat from TL-DNA from pTiB6S3"  
 FT misc\_feature 26..53  
 FT /tag= b  
 FT /note= "Synthetic polylinker derived sequences"  
 FT misc\_feature 54..90  
 FT /tag= c  
 FT /note= "Residual sequence from TL-DNA at right  
 FT border repeat"  
 FT misc\_feature 91..97  
 FT /tag= d  
 FT /note= "Synthetic polylinker derived sequences"  
 FT 3'UTR complement (98..305)  
 FT /tag= e  
 FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"  
 FT misc\_feature 310..330  
 FT /tag= f  
 FT /note= "Synthetic polylinker derived sequences"  
 FT CDS complement (331..882)  
 FT /tag= g  
 FT /note= "Streptomyces hygroscopicus bialaphos  
 FT resistance (bar) gene"  
 FT promoter complement (883..2608)  
 FT /tag= h  
 FT /note= "AtSIA ribulose-1,5-bisphosphate carboxylase  
 FT small subunit gene from Arabidopsis thaliana"  
 FT misc\_feature 2609..2658  
 FT /tag= i  
 FT /note= "Synthetic polylinker derived sequences"  
 FT

Search completed: February 25, 2002, 18:17:24  
Job time: 16682 sec

```

FT 3'UTR complement (2659..2919)
FT /*tag= j
FT /note= "taql fragment from 3' UTR of nopaline
FT synthase gene from T-DNA of pTiT37 and
FT containing plant polyadenylation signals"
FT misc_feature 2920..2940
FT /*tag= k
FT /note= "synthetic polylinker derived sequences"
FT 3'UTR 2941..2980
FT /*tag= l
FT /note= "downstream of Bacillus amyloliquefaciens
FT barstar coding region"
FT CDS complement (2981..3253)
FT /*tag= m
FT /note= "Barstar gene coding region from Bacillus
FT amyloliquefaciens"
FT promoter complement (3254..4762)
FT /*tag= n
FT /note= "anther-specific gene TA29 promoter from
FT Nicotiana tabacum"
FT misc_feature 4763..4807
FT /*tag= o
FT /note= "synthetic polylinker derived sequences"
FT misc_feature 4808..4832
FT /*tag= p
FT /note= "left border repeat from TL-DNA from pTiB6S3"
XX
XX WO200141558-A1.
XX
XX 14-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-EPI2872.
XX
XX 08-DEC-1999; 99US-0457037.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX De Both G, De Beuckeleer M;
XX
XX WPI; 2001-381419/40.
XX
XX Transgenic winter oilseed rape plants suited for producing hybrid seed
XX with improved qualities, comprises a male-sterility gene and fertility
XX restorer gene, integrated into the genome
XX
XX Example 1; Page 80-82; 98pp; English.
XX
XX The specification describes a pair of transgenic winter oilseed rape
XX plants suited for producing hybrid seed. One of the plants has an
XX expression cassette comprising a male-sterility gene, and the other
XX plant has an expression cassette comprising a fertility restorer gene,
XX integrated into the genome. The fertility restorer gene is capable of
XX preventing the activity of the male-sterility gene. The plant pair is
XX useful for producing hybrid seed. Plants developed from the hybrid
XX seed have agronomic performance, genetic stability and adaptability to
XX different genetic backgrounds. The present sequence represents
XX plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
XX a fertility restorer gene. The plasmid is used to create transgenic
XX plants of the invention.
XX
XX Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;
XX
Query Match 100.0%; Score 22; DB 22; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcattctacggcaatgtaccagc 22
Db 247 TCATCTACGGCAATGTACCAGC 226

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:15 ; Search time 301.6 seconds  
(without alignments)  
16.520 Million cell updates/sec

Title: US-09-698-903B-6  
Perfect score: 22  
Sequence: 1 tcattctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	1186	1 US-08-064-121-2	Sequence 2, Appli
2	22	100.0	1186	1 US-08-478-015-2	Sequence 2, Appli
3	22	100.0	1186	3 US-08-475-975-2	Sequence 2, Appli
4	22	100.0	1186	3 US-09-084-889-2	Sequence 2, Appli
5	22	100.0	1303	3 US-08-894-440-2	Sequence 2, Appli
6	22	100.0	3153	4 US-09-080-625-3	Sequence 3, Appli
7	22	100.0	3200	1 US-08-453-104-23	Sequence 23, Appli
8	22	100.0	3200	2 US-08-694-824-23	Sequence 23, Appli
9	22	100.0	3201	1 US-08-453-104-22	Sequence 22, Appli
10	22	100.0	3201	1 US-08-694-824-22	Sequence 22, Appli
11	22	100.0	3336	4 US-09-080-625-2	Sequence 2, Appli
12	22	100.0	3694	4 US-09-080-625-5	Sequence 5, Appli
13	22	100.0	3877	4 US-09-080-625-4	Sequence 4, Appli
c 14	22	100.0	4946	3 US-08-817-188-1	Sequence 1, Appli
c 15	22	100.0	5560	3 US-08-817-188-5	Sequence 5, Appli
c 16	22	100.0	5864	3 US-08-894-440-4	Sequence 4, Appli
c 17	22	100.0	5864	3 US-08-894-440-4	Sequence 4, Appli
c 18	22	100.0	6548	3 US-08-894-440-1	Sequence 1, Appli
c 19	22	100.0	6548	3 US-08-817-188-2	Sequence 2, Appli
c 20	22	100.0	7566	2 US-08-232-016-23	Sequence 23, Appli
c 21	22	100.0	7639	2 US-08-232-016-22	Sequence 22, Appli
c 22	22	100.0	7811	2 US-08-549-680A-5	Sequence 5, Appli
c 23	18.8	85.5	24595	6 5428147-1	Patent No. 5428147
c 24	16	72.7	29	2 US-08-232-016-12	Sequence 12, Appli
c 25	15.8	71.8	984	4 US-09-446-504-4	Sequence 4, Appli
c 26	15.8	71.8	3574	4 US-09-446-504-83	Sequence 83, Appli
c 27	15.8	71.8	3620	4 US-09-446-504-55	Sequence 55, Appli

c 28	15.2	69.1	238	1 US-07-914-284A-1	Sequence 1, Appli
c 29	15.2	69.1	238	5 PCT-US93-06645-1	Sequence 1, Appli
c 30	14.8	67.3	1290	4 US-09-247-373B-55	Sequence 55, Appli
c 31	14.8	67.3	1347	1 US-08-663-713A-1	Sequence 1, Appli
c 32	14.8	67.3	1347	3 US-09-014-888-1	Sequence 1, Appli
c 33	14.6	66.4	617	2 US-08-392-548C-3	Sequence 3, Appli
c 34	14.6	66.4	1512	4 US-08-955-918C-8	Sequence 8, Appli
c 35	14.6	66.4	1801	4 US-08-955-918C-6	Sequence 6, Appli
c 36	14.6	66.4	1804	1 US-08-631-200-1	Sequence 1, Appli
c 37	14.6	66.4	1804	1 US-08-829-553-1	Sequence 1, Appli
c 38	14.6	66.4	1804	2 US-08-922-267A-1	Sequence 1, Appli
c 39	14.6	66.4	1804	2 US-08-936-707A-1	Sequence 1, Appli
c 40	14.6	66.4	1804	2 US-08-936-706A-1	Sequence 1, Appli
c 41	14.6	66.4	1804	3 US-09-248-203-1	Sequence 1, Appli
c 42	14.6	66.4	1804	4 US-09-406-071-1	Sequence 1, Appli
c 43	14.6	66.4	2119	1 US-08-630-592-1	Sequence 1, Appli
c 44	14.6	66.4	2119	1 US-08-714-991-1	Sequence 1, Appli
c 45	14.6	66.4	2119	3 US-09-032-365A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-064-121-2  
; Sequence 2, Application US/08064121  
; Patent No. 5641664  
; GENERAL INFORMATION:  
; APPLICANT: D'HALLUIN, Kathleen  
; APPLICANT: GOBEL, Elke  
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/064,121  
; FILING DATE: 24-MAY-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 90403332.1  
; FILING DATE: 23-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 91401888.2  
; FILING DATE: 08-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crane-Feury, Sharon E  
; REGISTRATION NUMBER: 36,113  
; REFERENCE/DOCKET NUMBER: 010830-043  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1186 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: probe

```
;
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
;
US-08-064-121-2
```

```
Query Match 100.0%; Score 22; DB 1; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050
```

```
RESULT 2
US-08-478-015-2
; Sequence 2, Application US/08478015
; Patent No. 5712135
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,015
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 23-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2;
```

```
;
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
;
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine phosphotransferase ge
;
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylat
; OTHER INFORMATION: site derived from Agrobacterium T-DNA gene 7"
;
US-08-478-015-2
```

```
Query Match 100.0%; Score 22; DB 1; Length 1186;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
Db 1029 TCATCTACGGCAATGTACCAGC 1050
```

```
RESULT 3
US-08-475-975-2
; Sequence 2, Application US/08475975
; Patent No. 6002070
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,975
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/064,121
; FILING DATE: 24-MAY-1993
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
```



```
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-08-475-975-2
;
; Query Match 100.0%; Score 22; DB 3; Length 1186;
; Best Local Similarity 100.0%; Pred. No. 0.02;
; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcattctacggcaatgtaccagc 22
; |||||
; DB 1029 TCATCTACGGCAATGTACCAGC 1050
;
; RESULT 4
; US-09-084-889-2
; Sequence 2, Application US/09084889
; Patent No. 6074877
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,889
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/064,121
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA gene 7"
; US-09-084-889-2
;
; Query Match 100.0%; Score 22; DB 3; Length 1186;
; Best Local Similarity 100.0%; Pred. No. 0.02;
; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcattctacggcaatgtaccagc 22
; |||||
; DB 1029 TCATCTACGGCAATGTACCAGC 1050
;
; RESULT 5
; US-08-894-440-2
; Sequence 2, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1303
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI
; OTHER INFORMATION: fragment of pTS88
```

```
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(35)
;; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (35)..(694)
;; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain
;; OTHER INFORMATION: CM1841 (P35S)
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (695)..(967)
;; OTHER INFORMATION: region coding for barstar of Bacillus
;; OTHER INFORMATION: amyloliquefaciens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (968)..(1287)
;; OTHER INFORMATION: region containing polyadenylation signal of gene 7
;; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1288)..(1303)
;; OTHER INFORMATION: polylinker of pGEM2
US-08-894-440-2
```

```
Query Match 100.0%; Score 22; DB 3; Length 1303;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 tcattacggcaatgtaccagc 22
|||||
DB 1094 tcattacggcaatgtaccagc 1115
```

```
RESULT 6
US-09-080-625-3
; Sequence 3, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

```
; TOPOLOGY: linear
US-09-080-625-3

Query Match 100.0%; Score 22; DB 4; Length 3153;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22
|||||
DB 2675 TCATCTACGGCAATGTACCAGC 2696

RESULT 7
US-08-453-104-23
; Sequence 23, Application US/08453104
; Patent No. 5633446
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,104
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note="Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
US-08-453-104-23
```

```
Query Match 100.0%; Score 22; DB 1; Length 3200;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 tcattacggcaatgtaccagc 22  
 Db 2948 TCATCTACGCAATGTACCAGC 2969

RESULT 8  
 US-08-694-824-23  
 ; Sequence 23, Application US/08694824  
 ; Patent No. 5877306  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CORNELISSEN, Marc  
 ; APPLICANT: SOETAERT, Piet  
 ; APPLICANT: STAM, Maïke  
 ; APPLICANT: DOCKX, Jan  
 ; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
 ; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
 ; TITLE OF INVENTION: IN PLANT CELLS  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Burns, Doane, Swecker & Mathis  
 ; STREET: George Mason Bldg., Washington & Prince Sts.  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/694,824  
 ; FILING DATE: 09-AUG-1996  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/937,869  
 ; FILING DATE: 16-DEC-1992  
 ; APPLICATION NUMBER: GB 90401055.0  
 ; FILING DATE: 18-APR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rea, Teresa S  
 ; REGISTRATION NUMBER: 30,427  
 ; REFERENCE/DOCKET NUMBER: 010830-032  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 836-6620  
 ; TELEFAX: (703) 836-2021  
 ; INFORMATION FOR SEQ ID NO: 23:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3200 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: 2078..2082  
 ; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
 ; OTHER INFORMATION: wherein N is not known."  
 ; US-08-694-824-23

Query Match 100.0%; Score 22; DB 2; Length 3200;  
 Best Local Similarity 100.0%; Pred. No. 0.024;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
 Db 2948 TCATCTACGCAATGTACCAGC 2969

RESULT 9  
 US-08-453-104-22  
 ; Sequence 22, Application US/08453104

; Patent No. 5633446  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CORNELISSEN, Marc  
 ; APPLICANT: SOETAERT, Piet  
 ; APPLICANT: STAM, Maïke  
 ; APPLICANT: DOCKX, Jan  
 ; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
 ; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
 ; TITLE OF INVENTION: IN PLANT CELLS  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Burns, Doane, Swecker & Mathis  
 ; STREET: George Mason Bldg., Washington & Prince Sts.  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/453,104  
 ; FILING DATE:  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/937,869  
 ; FILING DATE: 16-DEC-1992  
 ; APPLICATION NUMBER: GB 90401055.0  
 ; FILING DATE: 18-APR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rea, Teresa S  
 ; REGISTRATION NUMBER: 30,427  
 ; REFERENCE/DOCKET NUMBER: 010830-032  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 836-6620  
 ; TELEFAX: (703) 836-2021  
 ; INFORMATION FOR SEQ ID NO: 22:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3201 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: 2151..2155  
 ; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
 ; OTHER INFORMATION: wherein N is not known."  
 ; US-08-453-104-22

Query Match 100.0%; Score 22; DB 1; Length 3201;  
 Best Local Similarity 100.0%; Pred. No. 0.024;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccagc 22  
 Db 3021 TCATCTACGCAATGTACCAGC 3042

RESULT 10  
 US-08-694-824-22  
 ; Sequence 22, Application US/08694824  
 ; Patent No. 5877306  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CORNELISSEN, Marc  
 ; APPLICANT: SOETAERT, Piet  
 ; APPLICANT: STAM, Maïke  
 ; APPLICANT: DOCKX, Jan  
 ; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
 ; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION

```
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 18-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 2151..2155
; OTHER INFORMATION: /note= "Nucleotides 2151-2155
; OTHER INFORMATION: wherein N is not known."
US-08-694-824-22

Query Match 100.0%; Score 22; DB 2; Length 3201;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22
Db 3021 TCATCTACGCGCAATGTACCAGC 3042

RESULT 11
US-09-080-625-2
; Sequence 2, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3336 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-080-625-2

Query Match 100.0%; Score 22; DB 4; Length 3336;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattcagcgcaatgtaccagc 22
Db 2858 TCATCTACGCGCAATGTACCAGC 2879

RESULT 12
US-09-080-625-5
; Sequence 5, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
; APPLICANT: Kriz, Alan L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-080-625-5
```

Query Match 100.0%; Score 22; DB 4; Length 3694;  
 Best Local Similarity 100.0%; Pred. No. 0.024; 0; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 DB 3216 TCATCTACGGCAATGTACCAGC 3237

RESULT 13  
 US-09-080-625-4  
 ; Sequence 4, Application US/09080625  
 ; Patent No. 6307123  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kriz, Alan L.  
 ; APPLICANT: Spencer, T. Michael  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE  
 ; TITLE OF INVENTION: IDENTIFICATION  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold White & Durkee  
 ; STREET: P.O. Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77210-4433  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/080,625  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hanson, Robert E.  
 ; REGISTRATION NUMBER: P-42,628  
 ; REFERENCE/DOCKET NUMBER: DEKM:161  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512) 418-3000  
 ; TELEFAX: (512) 474-7577  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3877 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-09-080-625-4

Query Match 100.0%; Score 22; DB 4; Length 3877;  
 Best Local Similarity 100.0%; Pred. No. 0.024;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 DB 3399 TCATCTACGGCAATGTACCAGC 3420

RESULT 14  
 US-08-817-188-1/c  
 ; Sequence 1, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366  
 ; EARLIER FILING DATE: 1996-07-31

; EARLIER APPLICATION NUMBER: EP 95401844.6  
 ; EARLIER FILING DATE: 1995-08-04  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 4946  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
 ; OTHER INFORMATION: plasmid pTHW107  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((1)..(25))  
 ; OTHER INFORMATION: T-DNA right border (RB)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((97)..(330))  
 ; OTHER INFORMATION: 3'g7: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
 ; OTHER INFORMATION: T-DNA  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((331)..(882))  
 ; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl  
 ; OTHER INFORMATION: transferase  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((883)..(2608))  
 ; OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
 ; OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((2658)..(3031))  
 ; OTHER INFORMATION: 3' nos: 3' untranslated region containing the  
 ; OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
 ; OTHER INFORMATION: gene of Agrobacterium T-DNA  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3032)..(3367))  
 ; OTHER INFORMATION: barnase: region coding for barnase  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((3368)..(4876))  
 ; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana  
 ; OTHER INFORMATION: tabacum  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: Complement((4922)..(4946))  
 ; OTHER INFORMATION: LB: T-DNA left border  
 ; US-08-817-188-1

Query Match 100.0%; Score 22; DB 3; Length 4946;  
 Best Local Similarity 100.0%; Pred. No. 0.025;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
 |||||  
 DB 247 TCATCTACGGCAATGTACCAGC 226

RESULT 15  
 US-08-817-188-5/c  
 ; Sequence 5, Application US/08817188  
 ; Patent No. 6074876  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DE BLOCK, MARC  
 ; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
 ; FILE REFERENCE: 2121-0127P  
 ; CURRENT APPLICATION NUMBER: US/08/817,188  
 ; CURRENT FILING DATE: 1997-05-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/03366

```
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTI1B6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uida: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-LS1 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5087)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTI1B6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
```

US-08-817-188-5

```
Query Match      100.0%; Score 22; DB 3; Length 5560;
Best Local Similarity 100.0%; Pred. No. 0.026;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22
   |||||
Db 234 TCATCTACGGCAATGTACCAGC 213
```

Search completed: February 25, 2002, 18:05:16  
Job time: 18564 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:01 ; Search time 8261.74 Seconds  
(without alignments)  
28.615 Million cell updates/sec

Title: US-09-698-903B-6  
Perfect score: 22  
Sequence: 1 tcatctacggcaatgtaccagc 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estfun:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estom:\*
- 5: em\_estpl:\*
- 6: em\_estba:\*
- 7: em\_estro:\*
- 8: em\_estov:\*
- 9: em\_htc:\*
- 10: gb\_est1:\*
- 11: gb\_est2:\*
- 12: gb\_htc:\*
- 13: gb\_gss:\*
- 14: em\_gss\_fun:\*
- 15: em\_gss\_hum:\*
- 16: em\_gss\_inv:\*
- 17: em\_gss\_pln:\*
- 18: em\_gss\_pro:\*
- 19: em\_gss\_rod:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					
c 1	17.4	79.1	568	11	B1174209
2	17.2	78.2	330	10	AW358852
3	17.2	78.2	352	10	AA475114
4	17.2	78.2	355	10	BE032643
5	17.2	78.2	355	11	BE924637
6	17.2	78.2	432	10	AA212655
7	17.2	78.2	562	13	AQ724420
8	17.2	78.2	567	13	AZ851447
9	17.2	78.2	721	11	BG592665
10	17.2	78.2	721	11	BG964419
c 11	17.2	78.2	932	11	BG400800
12	16.8	76.4	162	13	AZ474239

13	16.8	76.4	855	11	BG669143
c 14	16.4	74.5	479	10	AA458389
c 15	16.2	73.6	216	10	BE168752
c 16	16.2	73.6	309	13	TA105D12Q
17	16.2	73.6	314	13	AZ319004
18	16.2	73.6	343	10	AW789094
c 19	16.2	73.6	352	11	T69520
20	16.2	73.6	413	10	AA637160
21	16.2	73.6	423	10	AI593113
22	16.2	73.6	428	10	AI280109
c 23	16.2	73.6	435	11	T88735
24	16.2	73.6	438	10	AV668989
25	16.2	73.6	440	11	NG2954
26	16.2	73.6	443	10	AA878226
c 27	16.2	73.6	450	10	AA256224
28	16.2	73.6	450	10	AA574351
c 29	16.2	73.6	457	11	BG730070
c 30	16.2	73.6	495	10	AI448899
31	16.2	73.6	501	10	AI338197
c 32	16.2	73.6	502	11	BI234655
33	16.2	73.6	505	13	AZ720352
c 34	16.2	73.6	508	10	AI862985
c 35	16.2	73.6	544	10	AI533411
36	16.2	73.6	553	10	AI580125
c 37	16.2	73.6	571	10	BE367030
c 38	16.2	73.6	609	10	BE367017
c 39	16.2	73.6	611	11	W37106
40	16.2	73.6	622	11	BE819452
c 41	16.2	73.6	644	11	BG049365
c 42	16.2	73.6	658	10	AW953929
43	16.2	73.6	701	10	BE542678
44	16.2	73.6	739	11	BF177958
45	16.2	73.6	753	10	BE393004

ALIGNMENTS

RESULT	1
LOCUS	B1174209/c
DEFINITION	OSTF013F8_1 AD-wrmcDNA Caenorhabditis elegans cDNA similar to C46F11.2, mRNA sequence.
ACCESSION	B1174209
VERSION	B1174209.1
KEYWORDS	EST
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS	Reboul,J., Vaglio,P., Tzellas,N., Thierry-Mieg,N., Moore,T., Jackson,C., Shin-i,T., Kohara,Y., Thierry-Mieg,D., Thierry-Mieg,J., Lee,H., Hitti,J., Doucette-Stamm,L., Hartley,J.L., Temple,G.P., Brasch,M.A., Vandenhaute,J., Lamesch,P.E., Hill,D.E. and Vidal,M.
TITLE	Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in C. elegans
JOURNAL	Nat. Genet. 27 (3), 332-336 (2001)
MEDLINE	21135099
COMMENT	Contact: Reboul J, Vaglio P Marc Vidal Laboratory Dana Farber Cancer Institute 44 Binney Street, Boston, MA 02115, USA Tel: 617 632 5180 Fax: 617 632 2425 Email: Jerome.Reboul@dfci.harvard.edu Sequence tag of Gateway entry clones. The primers used were designed on the predicted protein encoding ORF. C. elegans ORFome cloning project; Contact jerome_reboul@dfci.harvard.edu or Philippe_Vaglio@dfci.harvard.edu POLYA-No.

FEATURES  
source Location/Qualifiers  
1. .568

/organism="Caenorhabditis elegans"  
 /strain="N2"  
 /db\_xref="taxon:6239"  
 /clone\_lib="AD-wrmcDNA"  
 /sex="Hermaphrodite and male"  
 /tissue\_type="whole animal"  
 /dev\_stage="mixed stage"  
 /note="The AD-wrmcDNA library was generated with poly(A)+  
 RNA isolated from both hermaphrodite and male N2 worms of  
 all larval stages, embryos, adults and dauers and the  
 subsequent generation of cDNAs by poly(A) priming. The  
 cDNAs were cloned into pPCR86"  
 146 a 116 c 159 g 147 t  
 BASE COUNT  
 ORIGIN

Query Match 79.1%; Score 17.4; DB 11; Length 568;  
 Best Local Similarity 94.7%; Pred. No. 3.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 tctacggcaatgtaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 564 TCTACGGCAATATACAGC 546

RESULT 2  
 AW358852  
 LOCUS AW358852 330 bp mRNA EST 09-JUL-2000  
 DEFINITION 43779 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION AW358852  
 VERSION AW358852.1 GI:6863502  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa

REFERENCE  
 AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 1 (bases 1 to 330)  
 Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,  
 Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.  
 and Keeler, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Smith TPL  
 USDA, ARS US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 20  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCCGAGTCACGACG  
 Plate: 25 row: A column: 16  
 Seq primer: ATTAGTGACACTAG.

FEATURES  
 source  
 Location/Qualifiers  
 1..330  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 2P1G"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
 Library made from pooled tissue from testis, ovary,  
 endometrium, hypothalamus, pituitary, and placenta."  
 54 a 118 c 108 g 50 t  
 BASE COUNT  
 ORIGIN

Query Match 78.2%; Score 17.2; DB 10; Length 330;  
 Best Local Similarity 86.4%; Pred. No. 4.2e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 tcatctacggcaatgtaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 190 TCATCCACGGCGTTGTACCAGC 211

RESULT 3  
 AA475114  
 LOCUS AA475114 352 bp mRNA EST 18-JUN-1997  
 DEFINITION v095a03.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA  
 clone IMAGE:873676 5', mRNA sequence.  
 ACCESSION AA475114  
 VERSION AA475114.1 GI:2199644  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 352)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

TITLE The WashU-HMNI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMNI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LML; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:513156  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 346.

FEATURES  
 source  
 Location/Qualifiers  
 1..352  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:873676"  
 /clone\_lib="Barstead mouse pooled organs MPLRB4"  
 /sex="mixed"  
 /tissue\_type="pooled organs"  
 /dev\_stage="7 day"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTTCAGATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 [GTTGATCGGTAC], digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library constructed by Bob Barstead."  
 57 a 118 c 101 g 75 t 1 others  
 BASE COUNT  
 ORIGIN

Query Match 78.2%; Score 17.2; DB 10; Length 352;  
 Best Local Similarity 86.4%; Pred. No. 4.2e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcatctacggcaatgtaccagc 22  
 ||||| ||||| ||||| |||||  
 Db 54 TCATCTACCGGATTGTACCAGC 75

RESULT 4  
 BE032643



polylinker; site\_1: Not I; site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCCGCGAATCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 71 a 148 c 125 g 88 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 10; Length 432;  
Best Local Similarity 86.4%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
||||||| | | | | | | | | | |

Db 94 TCATCTACCGGATGTACCAGC 115  
||||||| | | | | | | | | | |

RESULT 7  
A0724420 562 bp DNA GSS 14-JUL-1999  
LOCUS HS\_2119\_A2\_C07\_MR CIT Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone Plate=2119 Col=14 Row=E, DNA sequence.  
ACCESSION A0724420  
VERSION A0724420.1 GI:5484089  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and  
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 2119 row: E column: 14  
Seq primer: M13 Reverse  
Class: BAC ends

High quality sequence stop: 562.

Location/Qualifiers

1. .562  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate=2119 Col=14 Row=E"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 142 a 137 c 122 g 153 t 8 others

Query Match 78.2%; Score 17.2; DB 13; Length 562;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22  
||||||| | | | | | | | | | |

Db 438 TCATCTACGCCACTGTGCCAGC 459

RESULT 8  
A2851447/c

LOCUS A2851447

DEFINITION 2M0153M11R Mouse 10kb plasmid UUGCLM library Mus musculus genomic

clone UUGC2M0153M11 R, DNA sequence.

ACCESSION A2851447

VERSION A2851447.1 GI:13037453

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 567)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0153 row: M column: 11

Seq primer: CACACAGGAACACGATGACCC

Class: plasmid ends

High quality sequence stop: 567.

Location/Qualifiers

1. .567

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0153M11"

/clone\_lib="Mouse 10kb plasmid UUGCLM library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (g1473211419b/AP129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 188 a 94 c 106 g 179 t

Query Match 78.2%; Score 17.2; DB 13; Length 567;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattctacggcaatgtaccagc 22



5'-ATTCTAGAGCGCGAGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC Library." 212 a 397 c 131 q 192 t

BASE COUNT  
ORIGIN

Query Match 78.2%; Score 17.2; DB 11;  
Best Local Similarity 86.4%;  
Matches 19; Conservative 0; Mismatches 3; Indels 0;  
pred. No. 5.2e+02; Length 932;

**Oy** 1 tcaatctacggcgaatgtaccagc 22  
||| ||| ||| ||| ||| ||| ||| |||  
**Dd** 301 TGATCAGCGCAATGGACCAAC 280

RESULT 12  
AZ474239

LOCUS	AZ474239	162 bp	DNA
DEFINITION	M2090J011R Mouse 10kb plasmid UUCG1 library Mus musculus genomic clone UUCG1M2090J1 R, DNA sequence.		
ACCESSION	AZ474239		
VERSION	AZ474239.1	GI:10632364	
KEYWORDS	GSS		
SOURCE	house mouse,		
ORGANISM	Mus musculus		
			04-OCT-2000

REFERENCE  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 162)

**AUTHORS** Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb

**JOURNAL  
COMMENT**

Plasmid Inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0290 row: J column: 11  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 162.

FEATURES  
SOURCE

```

1: /invz
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="M0290311"
/clone_lib="Mouse 10kb plasmid UGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMW42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative

```

of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid  $\lambda$ 1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT	59 a
ORIGIN	

Query Match	76.4%;	Score	16.8;	DB	13;	Length	162;
Best Local Similarity	90.0%;	Pred. No.	5.6e+02;				
Matches	18;	Conservative	0;	Mismatches	2;	Indels	0

**Qy**      1    tcattctacggcaatgtacca 20  
         | | | | | | | | | |  
**Db**      43   TTATCTACTGCAATGTACCA 62

RESULT 13  
BG669143

LOCUS	BG669143	855 bp	mRNA	EST	30-APR-2001
DEFINITION	DRN05H06 Rat DRG Library norvegicus cDNA clone DRN05H06 5' mRNA sequence.				

ACCESSION BG669143  
 VERSION BG669143.1 GI:13891050  
 KEYWORDS EST.

SOURCE	Norway rat.
ORGANISM	<i>Rattus norv.</i>

Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa;  
Mammalia; Eutheria; Rodentia;  
Sciurognathi; Muridae; Murinae;

REFERENCE  
1 (bases 1 to 855)  
Rattus.

**AUTHORS**  
Xiao, H.S., Han, Z.G., Zhang, F.X., Huang, Q.H., Lu, Y.J., Bao, L., Fu, G.,  
Guo, C., Yan, Q., Jin, S.X., Zhu, Z.D., Xu, X.R., Li, N.G., Chen, Z. and  
Zhang, X.

**TITLE** Distinct gene expression profiles of rat dorsal root ganglion induced by peripheral nerve axotomy

JOURNAL  
COMMENT  
Unpublished (2001)  
Contact: Zhang Xu

**Laboratory of Sensory System  
Institute of Neuroscience  
320 Yue yang Road, Shanghai 200031, P.R.China  
Tel: 86-21-64748700-121  
Fax: 86-21-64711346**

This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)

PCR Primers  
FORWARD: T3

BACKWARD: T7  
Seq primer: T3

POLYA=NO

FEATURES	Location/Qualifiers
source	1 855

```

1: 833
organism="Rattus norvegicus"
strain="Sprague-Dawley"
accession="U116"
clone_id="DRNGSH06"
clone_lib="Rat DRG Library"
sex="male"
tissue_type="dorsal root ganglion"
dev_stage="adult"

```

BASE COUNT	240 a	190 c	174 g	246 t	5 others
ORIGIN					

Query Match	76.4%	Score 16.8;	DB 11;	Length 855;
Best Local Similarity	90.0%	Pred. No. 7.8e+02;		
Matches	18.	Conservative		

```

QY 3 atctacggcaatgtaccagc 22
||||| ||||||| |||||
Db 673 ATCTAGGGCAATGTCCACG 692

RESULT 14
AA458389/c
LOCUS AA458389 479 bp mRNA EST 06-JUN-1997
DEFINITION vq49f10.r1 Soares mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:864715 5' similar to gb:J03040 SPARC PRECURSOR (HUMAN);
gb:M20692 Mouse osteonectin (MOUSE);, mRNA sequence.
ACCESSION AA458389
VERSION AA458389.1 GI:2181109
KEYWORDS EST.
SOURCE Mus musculus
ORGANISM house mouse.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 479)
Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Thelsing,B., Wyllie,T., Lennon,G., Soares,B., Wilsson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:508803
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 441.
FEATURES
source
1. 479
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:864715"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTTCACCAATCTGAAGTGGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 114 a 125 c 127 g 112 t 1 others
ORIGIN

Query Match 74.5%; Score 16.4; DB 10; Length 479;
Best Local Similarity 94.4%; Pred. NO. 1.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtac 18
||||| ||||||| |||||
Db 312 TCATCCACGGCAATGTAC 295

us-09-698-903b-6.rst

```

Search completed: February 25, 2002, 17:21:04  
Job time: 16157 sec

```

RESULT 15
BE168752/c
LOCUS BE168752 216 bp mRNA EST 21-JUN-2000
DEFINITION QV1-HT0516-140300-107-e07 HT0516 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE168752
VERSION BE168752.1 GI:8631473
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 216)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=et2-QV1-HT0516-140
300-107-e07&t3=2000-03-14&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 216.
FEATURES
source
1. 216
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0516"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 42 a 75 c 59 g 40 t
ORIGIN

Query Match 73.6%; Score 16.2; DB 10; Length 216;
Best Local Similarity 85.7%; Pred. NO. 1.1e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcattacggcaatgtaccag 21
||||| ||||||| |||||
Db 46 TCAGCCACGGCAAGTACCAG 26

```

---





```

misc_feature 1..234
              /note="plant DNA"
misc_feature 235..415
              /note="r-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN
Query Match 100.0%; Score 415; DB 6; Length 415;
Best Local Similarity 100.0%; Pred. No. 2.2e-64;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgagttgggttcacgattttgggtttgactcttcaccattacattgaaactct 60
Db 1 gtcgagttgggttcacgattttgggtttgactcttcaccattacattgaaactct 60

QY 61 tacgtagagaacaactcacaagcatttaacatgcttcataataatgtacattac 120
Db 61 tacgtagagaacaactcacaagcatttaacatgcttcataataatgtacattac 120

QY 121 gtatatcacgctatacaaatagcagcaaaatccatgtaaacgagcgaggccacc 180
Db 121 gtatatcacgctatacaaatagcagcaaaatccatgtaaacgagcgaggccacc 180

QY 181 atggtttcaagtattataataattataataattatggttaggtacatggccgataa 240
Db 181 atggtttcaagtattataataattataataattatggttaggtacatggccgataa 240

QY 241 gaaagcgcaatttcagatgtaattccatcttcgaaagaataatagtttaaatattat 300
Db 241 gaaagcgcaatttcagatgtaattccatcttcgaaagaataatagtttaaatattat 300

QY 301 tgataaaatacaagtcagggtattattagtcgaagcaaaacataaaattattgatgcaag 360
Db 301 tgataaaatacaagtcagggtattattagtcgaagcaaaacataaaattattgatgcaag 360

QY 361 tttaattccgaatatttcataactgattatatacagctggtacattgacctag 415
Db 361 tttaattccgaatatttcataactgattatatacagctggtacattgacctag 415

RESULT 2
AX127757/c
LOCUS AX127757 416 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 10 from Patent WO0131042.
ACCESSION AX127757
VERSION AX127757.1 GI:14134404
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 416)
AUTHORS Weston, B. and de Heuckeleer, M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
1..416
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="3' border flanking region of elite event MS-B2"
misc_feature 1..193
misc_feature 194..416
              /note="r-DNA"
              /note="plant DNA"
BASE COUNT 137 a 72 c 54 g 152 t 1 others
ORIGIN

Query Match 91.8%; Score 380.8; DB 6; Length 416;
Best Local Similarity 98.5%; Pred. No. 2.5e-58;
Matches 404; Conservative 1; Mismatches 3; Indels 2; Gaps 2;

```

```

QY 7 ttgggtgctaatgattttgggtttgactcttcaccattacattgaaactcttacgga 66
Db 409 TTGGGTGTCATGATTTTGGGTTTGGTCTTCACTTCCACATTACATATTGAAACTCTTACGGA 350

QY 67 tgagaacaactcacaagcatttaacatgcttcataataatgtacattac-gtata 125
Db 349 TGAGNACAACTCACAAGCATTAAATCATGTTTCATATAATATATGATATATACGGGTATA 290

QY 126 tatacaogtatacaaaatagtcagcaaaatccatgtaaacgagcgaggccaccatggt 185
Db 289 TATACACGTATACAAATAGTAGTCAGCGAGAAATCCATGTAAACGACAGGGGCACCATGGT 230

QY 186 ttcaagttattataataattataataattatggttaggtacatggccgataaagaaaa 245
Db 229 TTC-AGTATTATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 171

QY 246 ggcaatttcagatgtaattccatcttcgaaagaataatagtttaaatattatttgata 305
Db 170 GGCAATTTGTAGATGTTAAATTCCTCTTGAAGAATAATAGTTTAAATATTATTTCGATA 111

QY 306 aatacaagtcagggtattattagtcgaagcaaaacataaaattatttgatgcaagtttaa 365
Db 110 AAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACATAAAWTTTATTGTATGCAAGTTTAA 51

QY 366 attcagaaatttcataactgattatatacagctggtacattgacctag 415
Db 50 ATTCAGAAATATTTCATAAATGATTATATATCAGCTGGTACATTGCCGTAG 1

RESULT 3
AX127748/c
LOCUS AX127748 5865 bp DNA PAT 15-MAY-2001
DEFINITION Sequence 1 from Patent WO0131042.
ACCESSION AX127748
VERSION AX127748.1 GI:14134395
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 5865)
AUTHORS Weston, B. and de Heuckeleer, M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
source
1..5865
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="T-DNA of plasmid pCO113"
BASE COUNT 1849 a 1095 c 1149 g 1772 t
ORIGIN

Query Match 46.1%; Score 191.4; DB 6; Length 5865;
Best Local Similarity 99.5%; Pred. No. 5.7e-25;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 223 gatgcatatggccgataaagaagcaattttagatgtaattcccatcttgaaagaaa 282
Db 5813 GATGTCATGTCGATPAGAAAGGCAATTTGTAGATGTTAATTCCTCATCTTTGAAAGAAA 5754

QY 283 tatagtttaaatattatttgataaaatacaagtcagggtatttatagtcacaaagcaaaaca 342
Db 5753 TATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACA 5694

QY 343 taaatttattgatgcaagtttaaatccagaataatttccaaactgattatatacagctgg 402
Db 5693 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATATCATGCTGG 5634

QY 403 tacattgacctag 415
Db 5633 TACATTGCCGTAG 5621

```

```
RESULT 4
AX063413/c
LOCUS AX063413 7599 bp DNA PAT 24-JAN-2001
DEFINITION Sequence 5 from Patent WO0100833.
ACCESSION AX063413
VERSION AX063413.1 GI:12541201
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
1 (bases 1 to 7599)
REFERENCE
AUTHORS Hofmann,B., Mollier,P. and Pelletier,G.
TITLE Promoter expressed specifically in the cells of plant roots,
recombinant vectors and host cells comprising same and transgenic
plants obtained
JOURNAL Patent: WO 0100833-A 5 04-JAN-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)
FEATURES
source
Location/Qualifiers
1..7599
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="ADN-T de PKB5"
BASE COUNT 1972 a 1938 c 1937 g 1752 t
ORIGIN
Query Match 46.1%; Score 191.4; DB 6; Length 7599;
Best Local Similarity 99.5%; Pred. No. 5.6e-25;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 223 gatgtacatgcccgaataaggaagcaattttagatgttaattcccatcttgaagaaa 282
DB 6771 GATGTACATGGTCGATAGAAAGGCAATTTCTAGATGTTAATTCCTTCCATCTTGAAGAAA 6712
QY 283 tatagttaaattattattgataaaatacaagaagtcaggtattattagtcacaaagcaaaaaca 342
DB 6711 TATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACA 6652
QY 343 taaattttattgatgcaagtttaaaattcagaataatttcaataactgattatcagctgg 402
DB 6651 TAAATTTATTGATGCAAGTTTAAATTTCAGAAATATTTCAATAACTGATTATATCAGCTGG 6592
QY 403 tacattgccgtag 415
DB 6591 TACATTGCCGTAG 6579
RESULT 5
E31991/c
LOCUS E31991 6539 bp DNA PAT 07-FEB-2001
DEFINITION Mutated barnase gene and transgenic plant thereof.
ACCESSION E31991
VERSION E31991.1 GI:13021588
KEYWORDS JP 2000041682-A/4.
SOURCE unidentified.
ORGANISM unclassified.
1 (bases 1 to 6539)
REFERENCE
AUTHORS Kazuyuki,H.F.N.N.
TITLE Mutated barnase gene and transgenic plant thereof
JOURNAL Patent: JP 2000041682-A 4 15-FEB-2000;
JAPAN TOBACCO INC
COMMENT
OS Escherichia coli LE392
PN JP 2000041682-A/4
PD 15-FEB-2000
PF 04-AUG-1998 JP 1998220060
PR
PI KAZUYUKI HAMADA,FUMIO NAKAKIDO
PC C12N15/09,A01H5/00,C12N5/10,C12N9/22/(C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00, (C12N5/00, C12R1:91)
CC
FH Key Location/Qualifiers
FT source 1..6539
/organism="Escherichia coli LE392".
FEATURES
Source
Location/Qualifiers
1..6539
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1755 a 1578 c 1519 g 1687 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6539;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatgcccgaataaggaagcaattttagatgttaattcccatcttgaagaaaat 283
DB 6533 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTCCATCTTGAAGAAAAT 6474
QY 284 atagttaaattattattgataaaatacaagaagtcaggtattattagtcacaaagcaaaaacat 343
DB 6473 ATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACAT 6414
QY 344 aaattttattgatgcaagtttaaaattcagaataatttcaataactgattatcagctggt 403
DB 6413 AAATTTATTGATGCAAGTTTAAATTTCAGAAATATTTCAATAACTGATTATATCAGCTGGT 6354
QY 404 acattgccgtag 415
DB 6353 ACATTGCCGTAG 6342
RESULT 6
A60109/c
LOCUS A60109 6548 bp DNA circular PAT 22-OCT-1999
DEFINITION Sequence 2 from Patent WO9706267.
ACCESSION A60109
VERSION A60109.1 GI:3715125
KEYWORDS
SOURCE Plasmid PTS172.
ORGANISM Plasmid PTS172
plasmids.
REFERENCE 1 (bases 1 to 6548)
AUTHORS De B.M.
TITLE Genetic transformation using a PAPP inhibitor
JOURNAL Patent: WO 9706267-A 2 20-FEB-1997;
PLANT GENETIC SYSTEMS NV (BE)
FEATURES
Source
Location/Qualifiers
1..6548
/organism="Plasmid PTS172"
/db_xref="taxon:106340"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatgcccgaataaggaagcaattttagatgttaattcccatcttgaagaaaat 283
DB 6542 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTCCATCTTGAAGAAAAT 6483
QY 284 atagttaaattattattgataaaatacaagaagtcaggtattattagtcacaaagcaaaaacat 343
DB 6482 ATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACAT 6423
QY 344 aaattttattgatgcaagtttaaaattcagaataatttcaataactgattatcagctggt 403
DB 6422 AAATTTATTGATGCAAGTTTAAATTTCAGAAATATTTCAATAACTGATTATATCAGCTGGT 6363
QY 404 acattgccgtag 415
|||||
```

```
Db 6362 ACATTGCCGTAG 6351
RESULT 7
LOCUS A76916 6548 bp DNA circular PAT 22-OCT-1999
DEFINITION Sequence 2 from Patent EP0757102.
ACCESSION A76916
VERSION A76916.1 GI:6088713
KEYWORDS
SOURCE
ORGANISM
Plasmid PTS172.
Plasmids.
1 (bases 1 to 6548)
REFERENCE
AUTHORS De,B.M.
TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR
JOURNAL Patent: EP 0757102-A 2 05-FEB-1997;
PLANT GENETIC SYSTEMS NV (BE)
FEATURES
Location/Qualifiers
source 1..6548
/organism="plasmid PTS172"
/db_xref="taxon:106340"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatggccgataaagaaggcaattttagatgtttaattcccatcttgaaagaat 283
Db 6542 ACGTACATGGTCGATAAGAAAGCAATTGTAGATGTTAATTCCTTGAAGAAAT 6483
QY 284 atagtttaaatattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCAGCAAAACAT 6423
QY 344 aaatttattgatgcaagtttaaatcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6363
QY 404 acattgccgtag 415
Db 6362 ACATTGCCGTAG 6351
RESULT 9
LOCUS E31990 6548 bp DNA PAT 07-FEB-2001
DEFINITION Mutated barnase gene and transgenic plant thereof.
ACCESSION E31990
VERSION E31990.1 GI:13021587
KEYWORDS JP 2000041682-A/3.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 6548)
AUTHORS Kazuyuki,H.F.N.N.
TITLE Mutated barnase gene and transgenic plant thereof
JOURNAL Patent: JP 2000041682-A 3 15-FEB-2000;
JAPAN TOBACCO INC
COMMENT OS Escherichia coli LE392
PN JP 2000041682-A/3
PD 15-FEB-2000
PF 04-AUG-1998 JP 1998220060
PR KAZUYUKI HAMADA,FUMIO NAKAKIDO
PC C12N15/09,A01H5/00,C12N5/10,C12N9/22/(C12N5/10,C12R1:91), PC
C12N15/00,
PC C12N5/00,(C12N5/00,C12R1:91)
CC
FH Key Location/Qualifiers
FT source 1..6548
/organism="Escherichia coli LE392".
FEATURES
Location/Qualifiers
source 1..6548
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1756 a 1579 c 1523 g 1690 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatggccgataaagaaggcaattttagatgtttaattcccatcttgaaagaat 283
Db 6542 ACGTACATGGTCGATAAGAAAGCAATTGTAGATGTTAATTCCTTGAAGAAAT 6483
QY 284 atagtttaaatattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCAGCAAAACAT 6423
QY 344 aaatttattgatgcaagtttaaatcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6363
QY 404 acattgccgtag 415
Db 6362 ACATTGCCGTAG 6351
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatggccgataaagaaggcaattttagatgtttaattcccatcttgaaagaat 283
Db 6542 ACGTACATGGTCGATAAGAAAGCAATTGTAGATGTTAATTCCTTGAAGAAAT 6483
QY 284 atagtttaaatattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCAGCAAAACAT 6423
QY 344 aaatttattgatgcaagtttaaatcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6363
QY 404 acattgccgtag 415
Db 6362 ACATTGCCGTAG 6351
Query Match 45.5%; Score 188.8; DB 6; Length 6548;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatggccgataaagaaggcaattttagatgtttaattcccatcttgaaagaat 283
```

```
RESULT 10
LOCUS AR078675 7811 bp DNA PAT 31-AUG-2000
DEFINITION Sequence 5 from patent US 5962768.
ACCESSION AR078675
VERSION AR078675.1 GI:10005421
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 7811)
AUTHORS Cornelissen,M., Reynaerts,A., Gosselle,V. and Van Aarssen,R.
TITLE Marker gene
JOURNAL Patent: US 5962768-A 5 05-OCT-1999;
FEATURES
source
BASE COUNT 1950 a 1906 c 1873 g 2082 t
ORIGIN
Query Match 45.5%; Score 188.8; DB 6; Length 7811;
Best Local Similarity 99.0%; Pred. No. 1.6e-24;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 224 atgtacatggccgataaagaaaggcaattgttagatgttaattcccatcttgaagaaat 283
DB 438 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAAT 497
QY 284 atagtttaaatattattatgataaaatacaagtcagggtatttagtcccaagcaaaacat 343
DB 498 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATTAGTCCCAAGCAAAACAT 557
QY 344 aaatttatgtcagtttaattcagaataatttcaataatttcaataactgattatcagctgt 403
DB 558 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGTT 617
QY 404 acattgcgctag 415
DB 618 ACATTGCCGTAG 629
RESULT 11
LOCUS I44104/c 3200 bp DNA PAT 07-OCT-1997
DEFINITION Sequence 23 from patent US 5633446.
ACCESSION I44104
VERSION I44104.1 GI:2469202
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 3200)
AUTHORS Cornelissen,M., Soetaert,P., Stam,M. and Dockx,J.
TITLE Modified Bacillus thuringiensis insecticidal-crystal protein genes
JOURNAL Patent: US 5633446-A 23 27-MAY-1997;
FEATURES
source
BASE COUNT 880 a 710 c 720 g 885 t 5 others
ORIGIN
Query Match 45.4%; Score 188.4; DB 6; Length 3200;
Best Local Similarity 99.5%; Pred. No. 2e-24;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 226 gtacatggccgataaagaaaggcaattgttagatgttaattcccatcttgaagaaat 285
DB 3141 GTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 3082
QY 286 agtttaaatattattatgataaaatacaagtcagggtatttagtcccaagcaaaacataa 345
```

```
Db 3081 AGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATTAGTCCCAAGCAAAACATAA 3022
QY 346 atttattgatgcaagtttaattcagaataatttcaataactgattatcagctggtac 405
Db 3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTTCANTACTGATTATTATCAGCTGGTAC 2962
QY 406 attgcgctag 415
Db 2961 ATTGCCGTAG 2952
RESULT 12
LOCUS AX172441 4832 bp DNA PAT 03-JUL-2001
DEFINITION Sequence 2 from Patent WO0141558.
ACCESSION AX172441
VERSION AX172441.1 GI:14597553
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4832)
AUTHORS de Both,G. and de Beuckeleer,M.
TITLE Hybrid winter oilseed rape and methods for producing same
JOURNAL Patent: WO 0141558-A 2 14-JUN-2001;
FEATURES
source
misc_feature 1528 a 883 c 932 g 1488 t 1 others
BASE COUNT
ORIGIN
Query Match 45.4%; Score 188.4; DB 6; Length 4832;
Best Local Similarity 99.5%; Pred. No. 2e-24;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 226 gtacatggccgataaagaaaggcaattgttagatgttaattcccatcttgaagaaat 285
Db 54 GTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 113
QY 286 agtttaaatattattatgataaaatacaagtcagggtatttagtcccaagcaaaacataa 345
Db 114 AGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATTAGTCCCAAGCAAAACATAA 173
QY 346 atttattgatgcaagtttaattcagaataatttcaataactgattatcagctggtac 405
Db 174 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTTCANTACTGATTATTATCAGCTGGTAC 233
QY 406 attgcgctag 415
Db 234 ATTGCCGTAG 243
RESULT 13
LOCUS A60108 4946 bp DNA PAT 06-MAR-1998
DEFINITION Sequence 1 from Patent WO9706267.
ACCESSION A60108
VERSION A60108.1 GI:3715124
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4946)
AUTHORS De.B.M.
TITLE Genetic transformation using a PARP inhibitor
JOURNAL Patent: WO 9706267-A 1 20-FEB-1997;
```

FEATURES  
source  
PLANT GENETIC SYSTEMS NV (BE)  
Location/Qualifiers  
1. .4946  
/organism="transformation vector pTHW107"  
/db\_xref="taxon:126810"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24; Mismatches 0; Indels 0; Gaps 0;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaaggaagcaattttagatgtttaattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGTCGATAGAAAAGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 113  
QY 286 agtttaaatattattgataaaataacaagtcaggtagtattattagtcacaagcaaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACATAA 173  
QY 346 attattgatgcagtttaataatcagaataatttcaataaactgattatcagctggtac 405  
|||||  
Db 174 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGTAC 233  
QY 406 attgcccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

RESULT 14  
LOCUS A76915 A76915 4946 bp DNA PAT 19-OCT-1999  
DEFINITION Sequence 1 from Patent EP0757102.  
ACCESSION A76915  
VERSION A76915.1 GI:6088712  
KEYWORDS  
SOURCE Transformation vector pTHW107.  
ORGANISM Transformation vector pTHW107.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De.B.M.  
TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR  
JOURNAL Patent: EP 0757102-A 1 05-FEB-1997;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES  
source  
Location/Qualifiers  
1. .4946  
/organism="transformation vector pTHW107"  
/db\_xref="taxon:126810"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24; Mismatches 0; Indels 0; Gaps 0;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaaggaagcaattttagatgtttaattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGTCGATAGAAAAGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 113  
QY 286 agtttaaatattattgataaaataacaagtcaggtagtattattagtcacaagcaaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACATAA 173  
QY 346 attattgatgcagtttaataatcagaataatttcaataaactgattatcagctggtac 405  
|||||  
Db 174 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGTAC 233  
QY 406 attgcccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

RESULT 15  
LOCUS AR098307 AR098307 4946 bp DNA PAT 14-FEB-2001  
DEFINITION Sequence 1 from patent US 6074876.  
ACCESSION AR098307  
VERSION AR098307.1 GI:12807564  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De Block,M.  
TITLE Genetic transformation using a PARP inhibitor  
JOURNAL Patent: US 6074876-A 1 13-JUN-2000;  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="unknown"  
BASE COUNT 1569 a 891 c 963 g 1523 t  
ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24; Mismatches 0; Indels 0; Gaps 0;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaaggaagcaattttagatgtttaattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGTCGATAGAAAAGCAATTGTAGATGTTAATCCCATCTTGAAGAATAAT 113  
QY 286 agtttaaatattattgataaaataacaagtcaggtagtattattagtcacaagcaaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACATAA 173  
QY 346 attattgatgcagtttaataatcagaataatttcaataaactgattatcagctggtac 405  
|||||  
Db 174 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGTAC 233  
QY 406 attgcccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

Search completed: February 25, 2002, 18:00:20  
Job time: 18428 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:17:24 ; Search time 716.55 seconds  
(without alignments)  
496.532 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagttgtgttcattga.....cagctggtacattgcgtag 415

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	415	100.0	415	22	Right (5') border
2	380.8	91.8	416	22	Left (3') border
3	191.4	46.1	5864	17	Plasmid pTCD113 T-
4	191.4	46.1	5865	22	Chimeric T-DNA of
5	191.4	46.1	7599	22	Nucleotide sequenc
6	188.8	45.5	5228	22	Plasmid pTSD172del
7	188.8	45.5	6539	21	E. coli plasmid pT
8	188.8	45.5	6548	17	Plasmid pTSD174 use
9	188.8	45.5	6548	18	Plasmid pTSD172. C
10	188.8	45.5	6548	21	E. coli plasmid pT
11	188.8	45.5	7492	22	Plasmid pTSD346. U

c 12	188.4	45.4	1303	17	AAT39337	Plasmid pTSD88 (Eco
c 13	188.4	45.4	3201	12	AAQ14529	pS029 Bt ICP codi
c 14	188.4	45.4	4832	22	AAH25423	Nucleotide sequenc
c 15	188.4	45.4	4946	18	AAT59531	T-DNA of plasmid p
c 16	188.4	45.4	4946	22	AAH25422	Nucleotide sequenc
c 17	188.4	45.4	5349	19	AAV23239	T-DNA of pTSD24.
c 18	188.4	45.4	5864	17	AAT39339	Plasmid pTCD113 T-
c 19	188.4	45.4	5865	22	AAH06990	Chimeric T-DNA of
c 20	188.4	45.4	7566	14	AAQ21160	Plasmid pS0212 co
c 21	188.4	45.4	7639	14	AAQ42159	Plasmid pTSD884 con
c 22	182.4	44.0	1037	11	AAQ04705	USP-Promoter-casse
c 23	182.4	44.0	1085	11	AAQ04703	Legumin-signalpept
c 24	182.4	44.0	1160	11	AAQ04706	USP-signalpeptide
c 25	179.4	43.2	1077	22	AAH25439	Right flanking reg
c 26	177	42.7	3201	12	AAQ15144	pV536 Bt ICP codin
c 27	153	36.9	1186	13	AAQ25707	Chimeric neo gene
c 28	146	35.2	3153	21	AAZ29122	Plasmid DV131 comp
c 29	146	35.2	3336	21	AAZ29121	Plasmid DV130 comp
c 30	146	35.2	3694	21	AAZ29124	Plasmid DV133 used
c 31	146	35.2	3877	21	AAZ29123	Plasmid DV132 used
c 32	146	35.2	24593	6	AAH50226	Sequence of opine
c 33	146	35.2	24596	6	AAH50182	Complete nucleotid
c 34	107.6	25.9	936	22	AAF58252	Oligonucleotide D1
c 35	107.6	25.9	936	22	AAF58254	Oligonucleotide D1
c 36	107.6	25.9	936	22	AAF58257	Oligonucleotide D1
c 37	107.6	25.9	936	22	AAF58259	Oligonucleotide D2
c 38	107.6	25.9	936	22	AAF58262	Oligonucleotide D2
c 39	107.6	25.9	938	22	AAF58255	Oligonucleotide D1
c 40	106.4	25.6	936	22	AAF58252	Oligonucleotide D1
c 41	106.4	25.6	936	22	AAF58254	Oligonucleotide D1
c 42	106.4	25.6	936	22	AAF58257	Oligonucleotide D1
c 43	106.4	25.6	936	22	AAF58259	Oligonucleotide D2
c 44	106.4	25.6	936	22	AAF58262	Oligonucleotide D2
c 45	106.4	25.6	938	22	AAF58255	Oligonucleotide D1

ALIGNMENTS

RESULT	1
AAAD06997	
ID	AAAD06997 standard; DNA; 415 BP.
XX	
AC	AAAD06997;
XX	
DT	
XX	06-AUG-2001 (first entry)
DE	Right (5') border flanking region of elite event MS-B2.
XX	
KW	MS-B2 elite event; transgenic Brassica plant; transformation event;
KW	male-sterility gene; ds.
XX	
OS	Chimeric - Agrobacterium sp.
OS	Chimeric - Brassica sp.
XX	
FH	Key Location/Qualifiers
FT	misc_feature 1..234
FT	/tag= a
FT	/note= "Corresponds to plant DNA"
FT	misc_feature 235..415
FT	/tag= b
FT	/note= "Corresponds to T-DNA"
XX	
PN	WO200131042-A2.
PD	03-MAY-2001.
XX	
PF	26-OCT-2000; 2000WO-EP10680.
XX	
PR	29-OCT-1999; 99US-0430497.
XX	
PA	(AVET ) AVENTIS CROPS SCIENCE NV.
XX	

PI Weston B, De Beuckeleer M;  
XX WPI; 2001-300517/31.  
XX  
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX  
XX Claim 11; Page 51; 53pp; English.  
PS  
CC The present invention relates to a transgenic Brassica plant or its  
CC seed, cells or tissues, characterised by harbouring a specific  
CC transformation event, particularly by the presence of a male-sterility  
CC gene, at a specific location in the Brassica genome. Transgenic  
CC Brassica plant is useful for producing a hybrid seed by crossing the  
CC transgenic plant with a male-fertile Brassica plant and harvesting the  
CC hybrid seed from the transgenic Brassica plant.  
CC The present sequence is right (5') border flanking region of elite event  
CC MS-B2.  
XX  
SQ Sequence 415 BP; 154 A; 55 C; 70 G; 136 T; 0 other;

Query Match 100.0%; Score 415; DB 22; Length 415;  
Best Local Similarity 100.0%; Pred. No. 3.5e-76;  
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtcgagtttggttcagtgatttgggttttgactcttcaccattacattgaactct 60  
Db 1 gtcgagtttggttcagtgatttgggttttgactcttcaccattacattgaactct 60  
QY 61 tacggatgagaacaactcacagcattaatcatgttccataataatatgtacattatc 120  
Db 61 tacggatgagaacaactcacagcattaatcatgttccataataatatgtacattatc 120  
QY 121 gtatatatacgcgtatatacaaatagtagcgaagaataatccatgtaagcagcgccacc 180  
Db 121 gtatatatacgcgtatatacaaatagtagcgaagaataatccatgtaagcagcgccacc 180  
QY 181 atggtttcaagtattataataattataataatttgtaggtggtacattgcccgtataa 240  
Db 181 atggtttcaagtattataataattataataatttgtaggtggtacattgcccgtataa 240  
QY 241 gaaagcgcaattttagatgttcaattcccatcttgaaagaataatagtttaaatatttat 300  
Db 241 gaaagcgcaattttagatgttcaattcccatcttgaaagaataatagtttaaatatttat 300  
QY 301 tgataaatacaagtcagggtattatagtcacagcaaaacataaatttattgacgaag 360  
Db 301 tgataaatacaagtcagggtattatagtcacagcaaaacataaatttattgacgaag 360  
QY 361 ttttaaatcagaataattttcaataactgatttatcatcagctgggtacattgcccgtag 415  
Db 361 ttttaaatcagaataattttcaataactgatttatcatcagctgggtacattgcccgtag 415

RESULT 2  
AAD06999/c  
ID AAD06999 standard; DNA; 416 BP.  
XX  
XX AAD06999;  
AC  
XX  
XX 06-AUG-2001 (first entry)  
DT  
XX  
XX Left (3') border flanking region of elite event MS-B2.  
DE  
XX  
XX MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; ds.  
KW  
XX  
XX Chimeric - Agrobacterium sp.  
OS  
XX  
XX Chimeric - Brassica sp.  
FH Key Location/Qualifiers

misc\_feature 1..193  
FT /\*tag= a  
FT /note= "Corresponds to T-DNA"  
FT 194..416  
FT /\*tag= b  
FT /note= "Corresponds to plant DNA"  
XX  
XX WO200131042-A2.  
PN  
XX  
XX 03-MAY-2001.  
PD  
XX  
XX 26-OCT-2000; 2000WO-EPI0680.  
PF  
XX  
XX 29-OCT-1999; 99US-0430497.  
PR  
XX  
XX (AVET ) AVENTIS CROPS SCIENCE NV.  
PA  
XX  
XX Weston B, De Beuckeleer M;  
PI  
XX  
XX WPI; 2001-300517/31.  
DR  
XX  
XX Transgenic Brassica plants, seeds, cells or tissues, characterized by  
PT harboring specific transformation events, particularly by presence of  
PT male-sterility gene, at specific location in its genome -  
XX  
XX Claim 11; Page 52; 53pp; English.  
PS  
XX  
XX The present invention relates to a transgenic Brassica plant or its  
CC seed, cells or tissues, characterised by harbouring a specific  
CC transformation event, particularly by the presence of a male-sterility  
CC gene, at a specific location in the Brassica genome. Transgenic  
CC Brassica plant is useful for producing a hybrid seed by crossing the  
CC transgenic plant with a male-fertile Brassica plant and harvesting the  
CC hybrid seed from the transgenic Brassica plant.  
CC The present sequence is left (3') border flanking region of elite event  
CC MS-B2.  
XX  
SQ Sequence 416 BP; 137 A; 72 C; 54 G; 152 T; 1 other;

Query Match 91.8%; Score 380.8; DB 22; Length 416;  
Best Local Similarity 98.5%; Pred. No. 3.2e-69;  
Matches 404; Conservative 1; Mismatches 3; Indels 2; Gaps 2;  
QY 7 ttggtgttccatgatttgggttttgactcttcaccattacattgaactcttcacgga 66  
Db 409 TTGGGTCTTCATGATTTGGGTTTGGACTCTTCACCATTTACATATTGAAACTCTTACGGA 350  
QY 67 tgagaacaactcacagcattaatcatgttccataataatatgtacattatc-gtata 125  
Db 349 TGAGAACAACTCACAGCATTAATCATGTTCATATAAATATATGTACATTATACGGTATA 290  
QY 126 tatacacgtatatacaaatagtagcgaagaataatccatgtaagcagcgccaccattggt 185  
Db 289 TATACAGTATACAAATAGTAGCGGAGAAATCATGTAAGCAGCAGGGCCACCATGCT 230  
QY 186 ttcaagtattataataattataataatttgtaggtggtacattgcccgtataaagaaaa 245  
Db 229 TTC-AGTATTATATAAATTATAATTATGTTAGGATGTACATGGTGGTAAAGAAAA 171  
QY 246 ggcaattttagatgttaattcccatcttgaaagaataatagtttaaatattattgata 305  
Db 170 GGCAATTTCGTAGATGTTAATTCCCATCTTCCCAAGAAATATAGTTAAATATTATTGATA 111  
QY 306 aaataacaagtcagggtattatagtcacagcaaaacataaatttattgacgaagtttaa 365  
Db 110 AAATAACAAGTCAGGTATTATAGTCCAGCAAAAAACATAAATTTATTGATGCAAGTTAA 51  
QY 366 attcagaataattttcaataactgatttatcatcagctgggtacattgcccgtag 415  
Db 50 ATTGAGAAATATTTCATTAACATGATTATATATCAGCTGGTACATTGCCGTAG 1

```

RESULT 3
AAT39339/c
XX ID AAT39339 standard; DNA; 5864 BP.
XX AC AAT39339;
XX DT 22-JAN-1997 (first entry)
XX XX Plasmid pTCOLL3 T-DNA used to obtain male sterile oilseed rape.
DE DE Plasmid pTCOLL3; male sterile; barnase; ribonuclease; barstar;
KW KW transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX XX Synthetic.
OS OS
XX Key Location/Qualifiers
FH misc_feature complement (1..25)
FT /*tag= a
FT /label= RB
FT /note= "right border of Agrobacterium T-DNA"
FT polyA_signal complement (98..330)
FT /*tag= b
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT CDS complement (331..882)
FT /*tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin
FT acetyltransferase"
FT complement (883..2608)
FT /*tag= d
FT /label= Pssu
FT /note= "promoter of Arabidopsis Rubisco small
FT subunit gene"
FT polyA_signal complement (2659..3031)
FT /*tag= e
FT /label= 3'nos
FT /note= "region containing polyA signal of nopaline
FT synthase gene of Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /*tag= f
FT /label= Barnase
FT /note= "Bacillus amyloliquefaciens barnase coding
FT region"
FT promoter complement (3368..4877)
FT /*tag= g
FT /label= PTA29
FT /note= "promoter of stamen-specific TA29 gene of
FT Nicotiana tabacum"
FT promoter 4924..5216
FT /*tag= h
FT /label= Pnos
FT /note= "promoter of nopaline synthase gene of
FT Agrobacterium T-DNA"
FT CDS 5217..5489
FT /*tag= i
FT /label= Barstar
FT /note= "region coding for barstar of Bacillus
FT amyloliquefaciens"
FT polyA_signal 5490..5765
FT /*tag= j
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT misc_feature complement (5840..5864)
FT /*tag= k
FT /label= LB
FT /note= "left border of Agrobacterium T-DNA"
XX
XX W09626283-A1.
XX
XX 29-AUG-1996.

```

---

```

XX 21-FEB-1996; 96WO-EP00722.
XX PF
XX 21-FEB-1995; 95EP-0400364.
XX PR
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PA
XX Botterman J, Cornelissen M, Michiels F;
XX PI
XX WPI; 1996-402373/40.
XX DR
XX Prodn. of male sterile plants by transforming with a chimaeric
XX construct - comprising a male sterility DNA e.g. barnase and a
XX co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX for generating hybrid cultivars
XX PT
XX Example 3; Page 33-3743-47; 56pp; English.
XX PS
XX Plasmid pTCOLL3 (AAT39339) is a T-DNA vector containing a bar gene
XX under control of the PSU promoter, a barnase gene under control
XX of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
XX gene under control of the Pnos promoter. 87% Of oilseed rape
XX plants regenerated after Agrobacterium-mediated transformation
XX using pTCOLL3 were male sterile. Barnase expression disturbed the
XX function of stamen cells leading to male sterility. Constitutive
XX expression of barstar counteracted any low level expression of
XX barnase in non-stamen tissue.
XX CC
XX Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;
XX SQ

```

---

```

Query Match 46.1%; Score 191.4; DB 17; Length 5864;
Best Local Similarity 99.5%; Pred. No. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 223 gatgtacatggccgataaagaagcaattgttagatgttaattcccatcttgaagaaa 282
Db 5812 GATGTACATGGTCGATAGAGAAAGGCAATTTGTAGATGTTAATTCCTCCATCTTGAAGAAA 5753
Qy 283 tatagtttaataattattgataaaataacaagtcagattattatagtcacagcaaaaaca 342
Db 5752 TATAGTTTAAATATTATTATTGATAAAATAACAAGTCAGGTATTTATAGTCCAGCAAAAAACA 5693
Qy 343 taaattattgatgcaagtttaaatcagaataattccaataactgattatatacagctgg 402
Db 5692 TAAATTTATGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATATCAGCTGG 5633
Qy 403 tacattgccgtag 415
Db 5632 TACATTGCCGTAG 5620

```

---

```

RESULT 4
AAD06990/c
XX ID AAD06990 standard; DNA; 5865 BP.
XX AC
XX AAD06990;
XX DT 06-AUG-2001 (first entry)
XX DE
XX Chimeric T-DNA of plasmid pTCOLL3.
XX T-DNA: plasmid pTCOLL3; transgenic Brassica plant; transformation event;
XX male-sterility gene; chimeric; tobacco; ds.
XX OS
XX Chimeric - Streptomyces hygroscopicus.
XX Chimeric - Arabidopsis thaliana.
XX Chimeric - Bacillus amyloliquefaciens.
XX Chimeric - Nicotiana tabacum.
XX Chimeric - Agrobacterium tumefaciens.
XX Chimeric - Unidentified.
XX FH Key Location/Qualifiers

```



```
Qy 403 tacattgccgtag 415
|||||
Db 5633 TACATTGCCGTAG 5621

RESULT 5
AAF25320/c
ID AAF25320 standard; DNA; 7599 BP.
XX
AC AAF25320;
XX
DT 30-APR-2001 (first entry)
XX
XX Nucleotide sequence of a plasmid pGKB5.
XX
XX Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Synthetic.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR01768.
XX
PR 25-JUN-1999; 99FR-0008185.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
XX WPI; 2001-102893/11.
XX
XX New constitutive plant promoter active specifically in roots, useful
PT for controlling expression of pest or drought resistance genes, and
PT related transgenic plants -
XX
XX Disclosure; Fig 9; 92pp; French.
XX
CC The present sequence represents a plasmid pGKB5. The plasmid contains
CC a plant promoter that directs expression of a selected sequence in
CC root cells at all stages of development of a plant. The plant promoter
CC is used to control expression of genes in a root-specific manner,
CC especially genes that provide resistance to parasites, pests (nematodes
CC or fungi), water and salt stress, or alter sugar content or nitrogen
CC transport. Fragments of the promoter are useful as probes or primers
CC to detect or amplify at least part of the promoter.
XX
SQ Sequence 7599 BP; 1972 A; 1938 C; 1937 G; 1752 T; 0 other;

Query Match 46.18; Score 191.4; DB 22; Length 7599;
Best Local Similarity 99.58; Pred. No. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 223 gatgtacatggccgataagaagaaggcaattgttagatgttaattcccatcttgaaagaa 282
|||||
Db 6771 GATGTACATGGTCCGATAGAAAAGGCAATTTGTAGATGTTAATTCCTTGAAGAAA 6712

Qy 283 tatagtttaaatattttatgataaaaataacaagtcaggtattatagtcacaagcaaaaaa 342
|||||
Db 6711 TATAGTTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACA 6652

Qy 343 taaatttattgatcgaagtttaaatccagaataatttcaataactgatttatcagctgg 402
|||||
Db 6651 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAAGTATTATATCAGCTGG 6592

Qy 403 tacattgccgtag 415
|||||
Db 6591 TACATTGCCGTAG 6579
```

```
RESULT 6
AAF86439/c
ID AAF86439 standard; DNA; 5228 BP.
XX
AC AAF86439;
XX
DT 25-JUN-2001 (first entry)
XX
XX Plasmid pTSL172delta.
XX
KW Male sterile plant; RNAase inhibitor; plasmid pTSL172delta; ds.
XX
OS Unidentified.
XX
PN WO200124616-A1.
XX
PD 12-APR-2001.
XX
XX 12-SEP-2000; 2000WO-JP06222.
XX
PR 30-SEP-1999; 99JP-0279307.
XX
PA (NISB ) JAPAN TOBACCO INC.
XX
XX Hamada K, Nakakido F;
XX
XX WPI; 2001-266212/27.
XX
XX Method for producing male sterile rice and maize by inserting RNase
PT gene and RNase inhibitor genes with promoters into the plant genome -
PT Disclosure; Page 14-17; 29pp; Japanese.
XX
XX The present invention relates to a method for producing male sterile
CC plants. The method comprises inserting a promoter fragment upstream of an
CC RNase gene and a second promoter, upstream of an RNase inhibitor protein
CC gene and inserting it into the plant genome. The method is useful for
CC producing male sterile tobacco, lettuce and rapeseed plants, but
CC preferably rice and maize. The present sequence is a vector used in
CC the method of the present invention.
XX
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;

Query Match 45.58; Score 188.8; DB 22; Length 5228;
Best Local Similarity 99.08; Pred. No. 4.6e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 224 atgtacatggccgataagaagaaggcaattgttagatgttaattcccatcttgaaagaaat 283
|||||
Db 5222 ACGTACATGGTTCGATAAGAAAAGGCAATTTGTAGATGTTAATTCCTTCAAGAAAT 5163

Qy 284 atagtttaaaattatttattgataaaaataacaagtcaggtattatagtcacaagcaaaaaa 343
|||||
Db 5162 ATAGTTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACAT 5103

Qy 344 aaatttattgatcgaagtttaaatccagaataatttcaataactgatttatcagctggt 403
|||||
Db 5102 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAAGTATTATATCAGCTGGT 5043

Qy 404 acattgccgtag 415
|||||
Db 5042 ACATTGCCGTAG 5031

RESULT 7
AAZ91097/c
ID AAZ91097 standard; DNA; 6539 BP.
XX
AC AAZ91097;
XX
DT 06-JUN-2000 (first entry)
```

XX DE E. coli plasmid pTS431 containing mutant barnase gene.  
XX KW Male sterile plant; mutant barnase gene; anther-specific expression;  
KW Low fidelity PCR; primer; plant breeding; ss.  
XX OS Synthetic.  
XX PN WO200008176-A1.  
XX PD 17-FEB-2000.  
XX PF 03-AUG-1999; 99WO-JP04167.  
XX PR 04-AUG-1998; 98JP-0220060.  
XX PA (NISR) JAPAN TOBACCO INC.  
XX PI Hamada K, Nakakido F;  
XX DR WPI; 2000-199581/17.  
XX PT Mutate barnase gene for efficient construction of plant transformants,  
PT particularly male sterile plants free from any undesirable characters  
PT by specifically expressing the gene alone in anther.  
XX PS Example 3; Page 23-27; 30pp; Japanese.  
XX CC The invention relates to the generation of male sterile plants by  
CC the introduction of a mutant barnase gene (AA291095) for expression  
CC specifically in the anther of a plant. This sequence represents the  
CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains  
CC the mutated barnase gene (AA291095) under control of the cauliflower  
CC mosaic virus 35S promoter. The vector also contains a region of the  
CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the  
CC barnase gene to plants via an Agrobacterium tumefaciens host cell.  
XX The transformed plant is used in plant breeding.  
XX SQ Sequence 6539 BP; 1755 A; 1578 C; 1519 G; 1687 T; 0 other;  
  
Query Match 45.5%; Score 188.8; DB 21; Length 6539;  
Best Local Similarity 99.0%; Pred. No. 4.7e-30;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 224 atgtacatgccgataagaagaagcaatttctgtatgttattccatcttgaagaagaat 283  
Db 6533 ACGTACATGGTCGATAAGAAAGCAATTTGTAGATGTTAATTCCTTGAAGCAAT 6474  
Qy 284 atagtttaaatattatttgataaataaacaagtcaggtattatagtcgaagcaaaacat 343  
Db 6473 ATAGTTTAAATATTATTGATAAATAACNAGTCAGGTATTATAGTCCAGCAAAACAT 6414  
Qy 344 aaatttattgacaggtttcaattcagaagaattttcaataactgattatcagctggt 403  
Db 6413 AAATTTATTGATCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGTT 6354  
Qy 404 acattgcccgtag 415  
Db 6353 ACATTCGCGTAG 6342  
  
RESULT 8  
AAT39336/C  
ID AAT39336 standard; DNA; 6548 BP.  
XX AC AAT39336;  
XX AC AAT39336;  
XX DT 22-JAN-1997 (first entry)  
XX DE Plasmid pTS174 used to obtain male sterile rice.  
XX KW Plasmid pTS174; male sterile; barnase; ribonuclease; transgenic plant;

KW Rice; Oryza sativa; ds; cyclic.  
XX Synthetic.  
XX OS  
XX FH Key Location/Qualifiers  
FT misc\_feature 1..2003  
FT /\*tag= a  
FT /label= Vector  
FT /note= "pUC19 derived vector sequences"  
FT complement (2019..2283)  
FT /\*tag= b  
FT /label= 3'nos  
FT /note= "region containing polyadenylation signal  
FT nopaline synthase gene of Agrobacterium  
FT T-DNA"  
FT complement (2284..2624)  
FT /\*tag= c  
FT /label= Barnase  
FT /product= Bacillus amyloliquefaciens barnase  
FT complement (2625..4313)  
FT /\*tag= d  
FT /label= PEI  
FT /function= promoter of the stamen-specific EI gene  
FT 4336..5710  
FT /\*tag= e  
FT /label= P35S  
FT /function= 35S promoter of cauliflower mosaic virus  
FT 5711..6262  
FT /\*tag= f  
FT /label= bar  
FT /product= phosphinothricin acetyltransferase  
FT 6263..6496  
FT /\*tag= g  
FT /label= 3'g7  
FT /function= region containing polyadenylation signal  
XX  
XX PN WO9626283-A1.  
XX PD 29-AUG-1996.  
XX PF 21-FEB-1996; 96WO-EP00722.  
XX PR 21-FEB-1995; 95EP-0400364.  
XX PA (PLBZ) PLANT GENETIC SYSTEMS NV.  
XX PI Botterman J, Cornelissen M, Michiels F;  
XX DR WPI; 1996-402373/40.  
XX prodn. of male sterile plants by transforming with a chimeric  
XX construct comprising a male sterility DNA e.g. barnase and a  
XX co-regulating gene, e.g. barstar, into the nuclear genome, useful  
XX for generating hybrid cultivars  
PS Example 1; Page 33-37; 56pp; English.  
XX Plasmid pTS174 (AAT39336) contains Bacillus barnase DNA under control  
XX of the stamen-specific PEI promoter. Embryogenic callus from rice  
XX cv. Kochihibiki was transformed with pTS174 alone or with pTS88  
XX (see also AAT39337), a plasmid contg. barstar DNA under control of a  
XX 35S promoter. With pTS174 alone, 1 male sterile line was recovered  
XX from 48 electroporation cuvettes. With both plasmids, 7 normal  
XX male sterile lines were recovered from 40 cuvettes. Barnase  
XX expression disturbed the function of stamen cells leading to male  
XX sterility. Constitutive expression of barstar counteracted any low  
XX level expression of barnase in non-stamen tissue.  
XX SQ Sequence 6548 BP; 1757 A; 1578 C; 1523 G; 1690 T; 0 other;  
  
Query Match 45.5%; Score 188.8; DB 17; Length 6548;

```
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 224 atgtacatgccgataaagaagcaattttagatgttaattcccatcttgaagaat 283
Db 6542 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTGAAGAAT 6483
Qy 284 atagtttaataattattgataaataacaagtcagggtattatagtcacaagcaaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACAT 6423
Qy 344 aaatttattgatgaagtttaattcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTACAGAAATATTTCAATAACTGATTATATCATCAGTGGT 6363
Qy 404 acattgccgtag 415
Db 6362 ACATTGCCGTAG 6351

RESULT 9
AAT61394/c
ID AAT61394 standard; DNA; 6548 BP.
XX AC AAT61394;
XX DT 07-MAY-1997 (first entry)
XX DE Plasmid pTS172.
XX KW Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PARP;
XX KW niacinamide; Agrobacterium; T-DNA; male sterile; barnase;
XX KW ribonuclease; RNase; cereal; wheat; Triticum aestivum;
XX KW plasmid pTS172; ds.
XX OS Chimeric Agrobacterium sp.;
XX OS Chimeric Oryza sativa;
XX OS Chimeric cauliflower mosaic virus.
XX FH Key Location/Qualifiers
FT 3'UTR complement (2019..2288)
FT /label= 3'nos
FT /note= "3' untranslated region contg. the poly-A
FT signal of Agrobacterium T-DNA nopaline
FT CDS complement (2289..2624)
FT /tag= b
FT /product= barnase
FT /tag= c
FT /label= PE1
FT /note= "promoter region of rice El gene"
FT promoter complement (4336..5710)
FT /tag= d
FT /label= P35S
FT /note= "35S promoter region of cauliflower mosaic
FT virus"
FT CDS 5711..6262
FT /tag= e
FT /label= Bar
FT /note= "phosphinothricin acetyltransferase"
FT 6243..6496
FT /tag= f
FT /label= 3'g7
FT /note= "3' untranslated region contg. the poly-A
FT signal of gene 7 of Agrobacterium T-DNA"
XX EP757102-A1.
XX 05-FEB-1997.
XX 04-AUG-1995; 95EP-0401844.

XX PR 04-AUG-1995; 95EP-0401844.
XX PA (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX PI De Block M;
XX WP1; 1997-111050/11.
XX prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
XX inhibitor - reduces the cultured cells response to stress and
XX reduces metabolism
XX PS Example 2; Page 17-20; 25pp; English.
XX CC Plasmid pTS172 (AAT61394) contains the barnase coding sequence under
XX control of the rice El gene stamen-specific promoter and a
XX phosphinothricin acetyltransferase coding sequence under control of
XX the CamV 35S promoter. Plasmid pTS172 and plasmid pTS772 (see also
XX AAT61395) were used to transform wheat Spring variety Pavon calli via
XX particle bombardment. Some calli were treated with the poly-(ADP-
XX ribose) polymerase inhibitor niacinamide before, or before and
XX after bombardment. Healthy, male sterile plants were regenerated
XX only from bombarded calli that were treated with niacinamide. This
XX was believed to be due to more faithful expression characteristics
XX of the integrated stamen-selective barnase gene in these calli
XX and regenerated shoots. For plants transformed with pTS172,
XX foreign DNA was stably incorporated in the wheat genome in 2-3
XX copies.
XX SQ Sequence 6548 BP; 1756 A; 1579 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 18; Length 6548;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 224 atgtacatgccgataaagaagcaattttagatgttaattcccatcttgaagaat 283
Db 6542 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTGAAGAAT 6483
Qy 284 atagtttaataattattgataaataacaagtcagggtattatagtcacaagcaaaaacat 343
Db 6482 ATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACAT 6423
Qy 344 aaatttattgatgaagtttaattcagaataatttcaataactgattatcagctggt 403
Db 6422 AAATTTATTGATGCAAGTTTAAATTACAGAAATATTTCAATAACTGATTATATCATCAGTGGT 6363
Qy 404 acattgccgtag 415
Db 6362 ACATTGCCGTAG 6351

RESULT 10
AAZ91096/c
ID AAZ91096 standard; DNA; 6548 BP.
XX AC AAZ91096;
XX DT 06-JUN-2000 (first entry)
XX DE E. coli plasmid pTS172 containing synthetic barnase gene.
XX KW Male sterile plant; mutant barnase gene; anther-specific expression;
XX KW low fidelity PCR; primer; plant breeding; ss.
XX OS Synthetic.
XX PN WO200008176-A1.
XX PD 17-FEB-2000.
XX
```

```

PF 03-AUG-1999; 99WO-JP04167.
XX
PR 04-AUG-1998; 98JP-0220060.
XX
PA (NISR) JAPAN TOBACCO INC.
XX
PI Hamada K, Nakakido F;
XX WPI; 2000-195581/17.
XX
PT Mutate barnase gene for efficient construction of plant transformants,
PT particularly male sterile plants free from any undesirable characters
PT by specifically expressing the gene alone in anther.
XX
PS Example 3; Page 19-23; 30pp; Japanese.
XX
CC The invention relates to the generation of male sterile plants by
CC the introduction of a mutant barnase gene (AAZ91095) for expression
CC specifically in the anther of a plant. This sequence represents the
CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains
CC the synthetic barnase gene (AAZ91094) under control of the cauliflower
CC mosaic virus 35S promoter. The vector also contains a region of the
CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the
CC barnase gene to plants via an Agrobacterium tumefaciens host cell.
CC The transformed plant is used in plant breeding.
XX
SQ Sequence 6548 BP; 1756 A; 1579 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6548;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggccgataagaagaaggcaattttagatgttaattcccatcttgaaagaaat 283
DB 6542 ACGTACATGGTTCGATAGAAAAGGCAATTGTAGATGTTAATTCCTTGAAGAAAT 6483
QY 284 atagttaaataattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
DB 6482 ATAGTTTAAATATTATTGATTAATAATACAGTCAGGTATTATAGTCCAAAGCAAAACAT 6423
QY 344 aaattattatgcagtttaattcagaataatttcaataactgattatcagctggt 403
DB 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATTTTCAATTAATGATATATCAGCTGGT 6363
QY 404 acattgcccgtag 415
DB 6362 ACATTGCCGTAG 6351

RESULT 11
AAF86441/c
ID AAF86441 standard; DNA; 7492 BP.
XX
AC AAF86441;
XX
DT 25-JUN-2001 (first entry)
XX
DE Plasmid pTS346.
XX
KW Male sterile plant; RNAase inhibitor; plasmid pTS346; ds.
XX
OS Unidentified.
XX
PN WO200124616-A1.
XX
PD 12-APR-2001.
XX
PF 12-SEP-2000; 2000WO-JP06222.
XX
PR 30-SEP-1999; 99JP-0279307.
XX
PA (NISR) JAPAN TOBACCO INC.

```

```

XX Hamada K, Nakakido F;
XX WPI; 2001-266212/27.
XX
PT Method for producing male sterile rice and maize by inserting RNase
PT gene and RNase inhibitor genes with promoters into the plant genome -
XX Disclosure; Page 19-23; 29pp; Japanese.
XX
CC The present invention relates to a method for producing male sterile
CC plants. The method comprises inserting a promoter fragment upstream of an
CC RNase gene and a second promoter, upstream of an RNase inhibitor protein
CC gene and inserting it into the plant genome. The method is useful for
CC producing male sterile tobacco, lettuce and rapeseed plants, but
CC preferably rice and maize. The present sequence is a vector used in
CC the method of the present invention.
XX
SQ Sequence 7492 BP; 1987 A; 1801 C; 1752 G; 1952 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 7492;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggccgataagaagaaggcaattttagatgttaattcccatcttgaaagaaat 283
DB 7486 ACGTACATGGTTCGATAGAAAAGGCAATTGTAGATGTTAATTCCTTGAAGAAAT 7427
QY 284 atagttaaataattattgataaaatacaagtcagggtattatagtcacagcaaaacat 343
DB 7426 ATAGTTTAAATATTATTGATTAATAATACAGTCAGGTATTATAGTCCAAAGCAAAACAT 7367
QY 344 aaattattatgcagttttaaattcagaataatttcaataactgattatcagctggt 403
DB 7365 AAATTTATTGATGCAAGTTTAAATTCAGAAATTTTCAATTAATGATATATCAGCTGGT 7307
QY 404 acattgcccgtag 415
DB 7306 ACATTGCCGTAG 7295

RESULT 12
AAT39337/c
ID AAT39337 standard; DNA; 1303 BP.
XX
AC AAT39337;
XX
DT 22-JAN-1997 (first entry)
XX
DE Plasmid pTS88 (EcoRI-HindIII fragment).
XX
KW Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..35
FT /*tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT promoter 36..694
FT /*tag= b
FT /label= p35S
FT /function= 35S promoter of cauliflower mosaic virus
FT CDS 695..967 strain CM1841
FT /*tag= c
FT /label= barstar
FT /product= Bacillus amyloliquefaciens barstar
FT polyA_signal 968..1287
FT /*tag= d

```



```
FT /label= 3'g7
FT /function= region containing polyadenylation signal
FT of gene 7 og Agrobacterium T-DNA
FT misc_feature 1288..1303
FT /*tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
XX WO9626283-A1.
XX
XX 29-AUG-1996.
XX
XX 21-FEB-1996; 96WO-EP00772.
XX
XX 21-FEB-1995; 95EP-0400364.
XX
XX (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
XX Botterman J, Cornelissen M, Michiels F;
XX
XX WPI; 1996-402373/40.
XX
XX Prodn. of male sterile plants by transforming with a chimaeric
XX construct - comprising a male sterility DNA e.g. barnase and a
XX co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX for generating hybrid cultivars
XX
XX Example 1; Page 38; 56pp; English.
XX
XX The HindIII-EcoRI fragment (AA139337) of plasmid pTS88 contains
XX barstar DNA under control of a 35S promoter. The plasmid was
XX used with pTS174 (see also AA139336) contg. barnase DNA under
XX control of the stamen-specific promoter E1 to produce male sterile
XX rice cv. Kochihibiki transgenic plants, and with plasmid pV136
XX (see also AA139338) contg. barnase DNA under control of the stamen-
XX specific PCA55 promoter to produce male sterile maize plants.
XX Expression of barnase (a ribonuclease) in the stamen leads to male
XX sterility. Constitutive expression of barstar counteracts possible
XX low level expression of barnase DNA in non-stamen tissue.
XX
XX Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;
XX
XX
XX Query Match 45.4%; Score 188.4; DB 17; Length 1303;
XX Best Local Similarity 99.5%; Pred. No. 5e-30;
XX Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 226 gtacatgccgataagaagaagcaatttctagatgttaattcccatcttgaagaataat 285
XX Db 1287 GTACATGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 1228
XX
XX QY 286 agtttaaatatttattgataaaataacaagtcaggtattatagtcgaagcaaaaacataa 345
XX Db 1227 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAGCAAAAACATAA 1168
XX
XX QY 346 atttattgatgcaagttaaattcagaataatttcaataactgattatcagctgggtac 405
XX Db 1167 ATTTATTGTCGAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGGTAC 1108
XX
XX QY 406 attgccgtag 415
XX Db 1107 ATTGCCGTAG 1098
XX
XX RESULT 13
XX AAQ14529/c
XX ID AAQ14529 standard; DNA; 3201 BP.
XX
XX AAQ14529;
XX
XX 27-JAN-1992 (first entry)
XX
XX PP5029 Bt ICP coding sequence.
```

```
XX
XX Bacillus thuringiensis; insecticidal crystal protein; ICP;
XX deletion; ss.
XX
XX Synthetic.
XX
XX WO9116432-A.
XX
XX 31-OCT-1991.
XX
XX 17-APR-1991; 91WO-EP007733.
XX
XX 18-APR-1990; 90EP-0401055.
XX
XX (PLAN-) PLANT GENETIC SYST.
XX
XX Cornelissen M, Soetaert P, Stam M, Dockx J;
XX
XX WPI; 1991-339820/46.
XX
XX Modified Bacillus thuringiensis insecticidal crystal protein
XX genes - having A and T sequences changed to G and C sequences
XX encoding same amino acids, for increased expression levels
XX
XX Disclosure; Fig 6(c); 78pp; English.
XX
XX "n" in the sequence refers to not known nucleotides.
XX
XX PP5029 is identical to pVE36 (AAQ15144), but carries both the amino-
XX terminal modification and the internal modification of the Bt ICP
XX coding sequence.
XX
XX See also AAQ14529, AAQ15142-44.
XX
XX Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;
XX
XX
XX Query Match 45.4%; Score 188.4; DB 12; Length 3201;
XX Best Local Similarity 99.5%; Pred. No. 5.3e-30;
XX Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 226 gtacatgccgataagaagaagcaatttctagatgttaattcccatcttgaagaataat 285
XX Db 3141 GTACATGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 3082
XX
XX QY 286 agtttaaatatttattgataaaataacaagtcaggtattatagtcgaagcaaaaacataa 345
XX Db 3081 AGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCAGCAAAAACATAA 3022
XX
XX QY 346 atttattgatgcaagttaaattcagaataatttcaataactgattatcagctgggtac 405
XX Db 3021 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATATCAGCTGGTAC 2962
XX
XX QY 406 attgccgtag 415
XX Db 2961 ATTGCCGTAG 2952
XX
XX RESULT 14
XX AAH25423
XX ID AAH25423 standard; DNA; 4832 BP.
XX
XX AC AAH25423;
XX
XX 22-AUG-2001 (first entry)
XX
XX Nucleotide sequence of plasmid pTHW118.
XX
XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX fertility restorer gene; barstar gene; ss.
XX
XX Synthetic.
XX
XX Streptomyces hygroscopicus.
XX
XX Arabidopsis thaliana.
XX
XX Bacillus amyloliquefaciens.
```

```

OS Nicotiana tabacum.
XX
FH Key Location/Qualifiers
FT misc_feature 1..25
FT /tag= a
FT /note= "right border repeat from TL-DNA from pTiB6S3"
FT 26..53
FT /tag= b
FT /note= "synthetic polylinker derived sequences"
FT 54..90
FT /tag= c
FT /note= "residual sequence from TL-DNA at right
FT border repeat"
FT 91..97
FT /tag= d
FT /note= "synthetic polylinker derived sequences"
FT complement (98...309)
FT /tag= e
FT /note= "3' UTR from TL-DNA gene 7 of pTiB6S3"
FT 310...330
FT /tag= f
FT /note= "synthetic polylinker derived sequences"
FT complement (331..882)
FT /tag= g
FT /note= "Streptomyces hygrosopicus bialaphos
FT resistance (bar) gene"
FT complement (883..2608)
FT /tag= h
FT /note= "ats1A ribulose-1,5-bisphosphate carboxylase
FT small subunit gene from Arabidopsis thaliana"
FT 2609..2658
FT /tag= i
FT /note= "synthetic polylinker derived sequences"
FT complement (2659..2919)
FT /tag= j
FT /note= "Tag1 fragment from 3' UTR of nopaline
FT synthase gene from T-DNA of pTiT37 and
FT containing plant polyadenylation signals"
FT 2920..2940
FT /tag= k
FT /note= "synthetic polylinker derived sequences"
FT 2941..2980
FT /tag= l
FT /note= "downstream of Bacillus amyloliquefaciens
FT barstar coding region"
FT complement (2981..3253)
FT /tag= m
FT /note= "Barstar gene coding region from Bacillus
FT amyloliquefaciens"
FT complement (3254..4762)
FT /tag= n
FT /note= "anther-specific gene TA29 promoter from
FT Nicotiana tabacum"
FT 4763..4807
FT /tag= o
FT /note= "synthetic polylinker derived sequences"
FT 4808..4832
FT /tag= p
FT /note= "left border repeat from TL-DNA from pTiB6S3"
XX
PN WO200141558-A1.
XX
XX 14-JUN-2001.
XX
XX 06-DEC-2000; 2000WO-EP12872.
XX
XX 08-DEC-1999; 99US-0457037.
XX
XX (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX De Both G, De Beuckeleer M;
XX
XX WPI; 2001-381419/40.

```

```

XX Transgenic winter oilseed rape plants suited for producing hybrid seed
PT with improved qualities, comprises a male-sterility gene and fertility
PT restorer gene, integrated into the genome -
XX
XX Example 1; Page 80-82; 98pp; English.
XX
CC The specification describes a pair of transgenic winter oilseed rape
CC plants suited for producing hybrid seed. One of the plants has an
CC expression cassette comprising a male-sterility gene, and the other
CC plant has an expression cassette comprising a fertility restorer gene,
CC integrated into the genome. The fertility restorer gene is capable of
CC preventing the activity of the male-sterility gene. The plant pair is
CC useful for producing hybrid seed. Plants developed from the hybrid
CC seed have agronomic performance, genetic stability and adaptability to
CC different genetic backgrounds. The present sequence represents
CC a plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
CC a fertility restorer gene. The plasmid is used to create transgenic
CC plants of the invention.
XX
SQ Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match 45.4%; Score 188.4; DB 22; Length 4832;
Best Local Similarity 99.5%; Pred. No. 5.5e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatggcgcgatagaagaagcaattttagatgttaattcccatcttgaagaataat 285
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 agtttaaatattattgataaaataacaaagtcaggtattatagtcacagcaaaacataa 345
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 114 agtttaaatattattgataaaataacaaagtcaggtattatagtcacagcaaaacataa 173
QY 346 atttattgacgaagtttaaatcagaataatttcaataactgattatcagctggtac 405
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 174 atttattgacgaagtttaaatcagaataatttcaataactgattatcagctggtac 233
QY 406 attgcccgtag 415
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
AAT59531
ID AAT59531 standard; DNA; 4946 BP.
AC AAT59531;
XX
XX 07-MAY-1997 (first entry)
XX
XX T-DNA of plasmid pTHW107.
XX
XX Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PARP;
KW nicinamide; Agrobacterium; T-DNA; male sterile; barnase;
KW ribonuclease; RNase; cereal; wheat; oilseed rape; Brassica napus;
KW plasmid pTHW107; ds.
XX
XX Chimeric Agrobacterium sp.;
OS Chimeric Arabidopsis thaliana;
OS Chimeric Nicotiana tabacum.
XX
XX Key Location/Qualifiers
FT misc_RNA complement (1..25)
FT /tag= a
FT /label= RB
FT /note= "T-DNA right border"
FT 3'UTR complement (97..330)
FT /tag= b
FT /label= 3'g7
FT /note= "3' untranslated region contg. the poly-A
FT signal of gene-7 of Agrobacterium T-DNA"

```

```
FT CDS complement (331..882)
FT /*tag= c
FT /label= Bar
FT /product= phosphinothricin acetyltransferase
FT promoter complement (883..2608)
FT /*tag= d
FT /label= PSSU
FT /note= "promoter region of Rubisco small subunit
FT gene of Arabidopsis thaliana"
FT 3'UTR complement (2658..3031)
FT /*tag= e
FT /label= 3'nos
FT /note= "3'untranslated region contg. the poly-A
FT signal of the nopaline-synthase gene of
FT Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /*tag= f
FT /label= Barnase
FT /product= barnase
FT promoter complement (3368..4876)
FT /*tag= g
FT /label= PTA29
FT /note= "promoter region of tobacco TA29 gene"
FT misc_RNA complement (4822..4946)
FT /*tag= h
FT /label= LB
FT /note= "T-DNA left border"
XX
XX EP757102-A1.
XX
XX 05-FEB-1997.
XX
XX 04-AUG-1995; 95EP-0401844.
XX
XX 04-AUG-1995; 95EP-0401844.
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX De Block M;
XX
XX WPI; 1997-111050/11.
XX
XX Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
XX inhibitor - reduces the cultured cells response to stress and
XX reduces metabolism
XX
XX Example 3; Page 13-16; 25pp; English.
XX
XX Plasmid pTHW107 is a vector carrying T-DNA (AAT59531) comprising a
XX barnase coding sequence under control of the tobacco TA29 gene
XX stamen-specific promoter and a phosphinothricin acetyltransferase
XX coding sequence under control of an Arabidopsis Rubisco small
XX subunit gene promoter. Oilseed rape hypocotyl explants were
XX infected with Agrobacterium tumefaciens C58ClRif carrying vector
XX pTHW107 and helper Ti plasmid pMP60. In some treatments, the
XX hypocotyls were treated with the poly-(ADP-ribose) polymerase
XX inhibitor niacinamide (250 mg/l) 4 days prior to infection.
XX Plants regenerated from niacinamide-treated transformed calli
XX had a low copy number and displayed less variation in the
XX expression profile of the transgenes.
XX
XX Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 45.4%; Score 188.4; DB 18; Length 4946;
Best Local Similarity 99.5%; Pred. No. 5.5e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtcacatggccgataagaagaagcaatttgtagatgtaattcccatcttgaagaataat 285
DB ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 54 gtcacatggtcgataagaaggcaatttgtagatgtaattcccatcttgaagaataat 113
QY 286 agtttaaatatttattgataaaaataacaagtcagggtattattagtcaccaagcaaaacataa 345
```

```
Db 114 agtttaaatatttattgataaaaataacaagtcagggtattattagtcaccaagcaaaacataa 173
QY 346 atttattgatgcaagttttataattcagaaaattttcaataaactgatttatcagctcgtgtac 405
DB 174 atttattgatgcaagttttataattcagaaaattttcaataaactgatttatcagctcgtgtac 233
QY 406 attgccgtag 415
DB 234 attgccgtag 243
```

Search completed: February 25, 2002, 18:17:34  
Job time: 16692 sec

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:16 ; Search time 301.6 seconds  
(without alignments)  
311.633 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagtggtgttcattga.....cagctgtgattacgcgtag 415

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	191.4	46.1	5864	3	US-08-894-440-4
C 2	188.8	45.5	6548	3	US-08-894-440-1
C 3	188.8	45.5	6548	3	US-08-817-188-2
C 4	188.8	45.5	7811	2	US-08-549-680A-5
C 5	188.4	45.4	1303	3	US-08-894-440-2
C 6	188.4	45.4	3200	1	US-08-453-104-23
C 7	188.4	45.4	3200	1	US-08-694-824-23
C 8	188.4	45.4	4946	3	US-08-817-188-1
C 9	188.4	45.4	5560	3	US-08-817-188-5
C 10	188.4	45.4	7566	2	US-08-232-016-23
C 11	188.4	45.4	7639	2	US-08-232-016-22
C 12	186.8	45.0	5864	3	US-08-894-440-4
C 13	177	42.7	3201	1	US-08-453-104-22
C 14	177	42.7	3201	2	US-08-694-824-22
C 15	153	36.9	1186	1	US-08-064-121-2
C 16	153	36.9	1186	1	US-08-478-015-2
C 17	153	36.9	1186	3	US-08-475-975-2
C 18	153	36.9	1186	3	US-09-084-889-2
C 19	146	35.2	3153	4	US-09-080-625-3
C 20	146	35.2	3336	4	US-09-080-625-2
C 21	146	35.2	3694	4	US-09-080-625-5
C 22	146	35.2	3877	4	US-09-080-625-4
C 23	143	34.5	24595	6	5428147-1
C 24	44.8	10.8	8654	1	US-08-920-812-6
C 25	44.8	10.8	8654	1	US-08-920-827-6
C 26	44.8	10.8	8654	1	US-08-921-177-6
C 27	44.8	10.8	8654	1	US-08-362-577C-6

28	44.8	10.8	8654	2	US-08-920-828-6	Sequence 6, Appli
C 29	44.6	10.7	5526	3	US-08-751-359-21	Sequence 21, Appl
C 30	43.6	10.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 31	43	10.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 32	42.6	10.3	8920	2	US-08-446-855A-1	Sequence 1, Appli
C 33	42.6	10.3	8920	4	US-09-150-741-1	Sequence 1, Appli
C 34	42	10.1	1316	2	US-08-871-924A-1	Sequence 1, Appli
C 35	40.4	9.7	1588	3	US-09-058-489-45	Sequence 45, Appli
C 36	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appli
C 37	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appli
C 38	39.8	9.6	24595	6	5428147-1	Patent No. 5428147
C 39	39.4	9.5	80595	4	US-09-078-294-3	Sequence 3, Appli
C 40	39.2	9.4	658	4	US-08-998-416-595	Sequence 595, App
C 41	38.8	9.3	636	4	US-08-998-416-1137	Sequence 1137, Ap
C 42	38.8	9.3	660	1	US-07-991-867B-32	Sequence 32, Appl
C 43	38.8	9.3	660	1	US-08-107-755A-32	Sequence 32, Appl
C 44	38.8	9.3	660	2	US-08-544-332-32	Sequence 32, Appl
C 45	38.8	9.3	1511	1	US-07-991-867B-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-08-894-440-4/c  
; Sequence 4, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894.440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 5864  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; FEATURE:  
; OTHER INFORMATION: plasmid pTCO113  
; NAME/KEY: misc.feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: Complement((98)..(330))  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: region coding for phosphinothricin acetyl  
; OTHER INFORMATION: transferase (bar)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of  
; OTHER INFORMATION: Arabidopsis (Pssu)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: Complement((2659)..(3031))  
; OTHER INFORMATION: region containing polyadenylation signal of  
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA  
; OTHER INFORMATION: (3'nos)  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: Complement((3032)..(3367))  
; OTHER INFORMATION: region coding for barnase of Bacillus  
; OTHER INFORMATION: amyloliquefaciens  
; FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Phos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4

Query Match          46.1%; Score 191.4; DB 3; Length 5864;
Best Local Similarity 99.5%; Pred.No. 2.7e-37;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps

QY 223 gatgcacatgccgcgataagaaaaaggccaattgtagatgttaattcccatcttgaagaaa 282
Db 5812 GATGTACATGCTGCATGAAGAAAAGGCAATTTCTAGATCTTAATCCCATCTTGAAAGAAA 5753
QY 283 tatagtttaaatatttattgataaaatacaaaagtcaggatcaggtatattagtcacaaacaa 342
Db 5752 TATAGTTTAAATATTATTGATGATAAATAACAAGTCAGGTATTATTAGTCCAGCAAAAACA 5693
QY 343 taaattattgatgcgaagtttaaaattcagaaataattccaataactgattatcatcagctgg 402
Db 5692 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTCCAATACTGATTATATCAGCTGG 5633
QY 403 tacattgccgtag 415
Db 5632 TACATTGCCGCTAG 5620

RESULT 2
US-08-894-440-1/c
; Sequence 1, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NMSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: plasmid pTrs174
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2003)
; OTHER INFORMATION: pUC19 derived vector sequences (vector)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2019)..(2283))
; OTHER INFORMATION: 3' nos: region containing polyadenylation signal

```

```

; LOCATION: Complement((2019)..(2288))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2289)..(2624))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: PE1: promoter region of E1 gene of rice
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5170)
; OTHER INFORMATION: P35S: 35S promoter region of Cauliflower Mosaic
; OTHER INFORMATION: Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5711)..(6262))
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; US-08-817-188-2

```

[illegible]

```

4
RESULT
US-08-549-680A-5
Sequence 5, Application US/08549680A
Patent No. 5962768
GENERAL INFORMATION:
APPLICANT: CORNELISSEN, MARCUS
APPLICANT: REYNAERTS, ARLETTE
APPLICANT: GOSSELE, VERONIQUE
APPLICANT: VAN AARSSEN, ROEL
TITLE OF INVENTION: MARKER GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/549,680A
> FILING DATE: 16 JANUARY 1996
> CLASSIFICATION: 800
> ATTORNEY/AGENT INFORMATION:
> NAME: SVENSSON, LEONARD R.
> REGISTRATION NUMBER: 30,330
> REFERENCE/DOCKET NUMBER: 2121-0111P
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (703) 205-8000
> TELEFAX: (703) 205-8050
> TELEX: 248345
> INFORMATION FOR SEQ ID NO: 5:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 7811 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: double
> TOPOLOGY: circular
> MOLECULE TYPE: DNA (synthetic)
> FEATURE:
> NAME/KEY: misc_recomb
> LOCATION: 1..7811
> OTHER INFORMATION: /label= vector pTRVA3
> FEATURE:
> NAME/KEY: misc_feature
> LOCATION: 194..218
> OTHER INFORMATION: /note= "T-DNA right border"
> FEATURE:
> NAME/KEY: misc_feature
> LOCATION: 484..684
> OTHER INFORMATION: /note= "the 3' end formation and
> OTHER INFORMATION: polyadenylation region of T-DNA
> FEATURE:
> NAME/KEY: CDS
> LOCATION: complement (729..1340)
> OTHER INFORMATION: /note= "the aac(6') coding
> OTHER INFORMATION: sequence"
> FEATURE:
> NAME/KEY: promoter
> LOCATION: 1341..1756
> OTHER INFORMATION: /label= 35S promoter
> FEATURE:
> NAME/KEY: misc_feature
> LOCATION: 3001..3023
> OTHER INFORMATION: /note= "T-DNA left border
> OTHER INFORMATION: sequences"
> US-08-549-680A-5

```

[illegible]

US-08-894-440-2/c

; Sequence 2, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894,440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-EcoRI  
; OTHER INFORMATION: fragment of pTS88  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; FEATURE:  
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (P35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amyloliquefians  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
US-08-894-440-2

Query Match 45.4%; Score 188.4; DB 3; Length 1303;  
Best Local Similarity 99.5%; Pred. No. 1.1e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgcccgcgataaagaagcgaatttctagatgttaattcccatcttgaaagaatat 285  
Db 1287 GTACATGTCGTAAGAAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAATAAT 1228  
Qy 286 agtttaaatatttattgataaaaatacaagtcagggtattatagtcacaaagcaaaacataa 345  
Db 1227 AGTTTAAATATTATTGATAAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACATAA 1168  
Qy 346 atttattgacgaagtttaattcaaatatttcaataactgattatatacagctggtac 405  
Db 1167 ATTTATTGATGCAAGTTTAAATTTCAGAAATATTTCAATAACTGATTATATCAGCTGGTAC 1108  
Qy 406 attgcccgtag 415  
Db 1107 ATTGCCGTAG 1098

RESULT 6

US-08-453-104-23/c  
; Sequence 23, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETABERT, Piet  
; APPLICANT: STAM, Maïke

; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2078..2082  
; OTHER INFORMATION: /note= "Nucleotides 2078-2082  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-23

Query Match 45.4%; Score 188.4; DB 1; Length 3200;  
Best Local Similarity 99.5%; Pred. No. 1.3e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgcccgcgataaagaagcgaatttctagatgttaattcccatcttgaaagaatat 285  
Db 3141 GTACATGTCGTAAGAAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAATAAT 3082  
Qy 286 agtttaaatatttattgataaaaatacaagtcagggtattatagtcacaaagcaaaacataa 345  
Db 3081 AGTTTAAATATTATTGATAAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAACATAA 3022  
Qy 346 atttattgacgaagtttaattcaaatatttcaataactgattatatacagctggtac 405  
Db 3021 ATTTATTGATGCAAGTTTAAATTTCAGAAATATTTCAATAACTGATTATATCAGCTGGTAC 2962  
Qy 406 attgcccgtag 415  
Db 2961 ATTGCCGTAG 2952

RESULT 7

US-08-694-824-23/c  
; Sequence 23, Application US/08694824  
; Patent No. 587306



```

; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION
; TITLE OF INVENTION: IN PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/694,824
; FILING DATE: 09-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,869
; FILING DATE: 16-DEC-1992
; APPLICATION NUMBER: GB 90401055.0
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Rea, Teresa S
; REGISTRATION NUMBER: 30,427
; REFERENCE/DOCKET NUMBER: 010830-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2078..2082
; OTHER INFORMATION: /note= "Nucleotides 2078-2082
; OTHER INFORMATION: wherein N is not known."
US-08-694-824-23

```

```

Query Match      45.4%; Score 188.4; DB 2; Length 3200;
Best Local Similarity 99.5%; Pred. No. 1.3e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgccataagaagcaattgttagatgtaattcccatcttgaagaataat 285
Db 3141 GTACATGTCGATAGAAAAGCGCAATTGTAGATGTTAATTCCTCCCATCTTGAAGAAATAT 3082

Qy 286 agtttaaatattattgataaaataacaagtcaggtattatagtcacagcaaaaacataa 345
Db 3081 AGTTTAAATATTATTGATATAAAATAACAAGTCAGGTATTATAGTCCAGCAAAAACATAA 3022

Qy 346 atttattgatgcaagttaaattcagaataatttcataaactgattatatacagctggtac 405
Db 3021 ATTATTATGATCAAGTTTAAATTCAGAAATATTTCAATAAAGTATTATATCAGCTGCTGAC 2962

Qy 406 attgccgtag 415
Db 2961 ATTGCCGTAG 2952

```

```

RESULT 8
US-08-817-188-1
; Sequence 1, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4946
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW107
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: T-DNA right border (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((97)..(330))
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: bar: region coding for phosphinotricin acetyl
; OTHER INFORMATION: transferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter region of Rubisco small subunit gene of
; OTHER INFORMATION: Arabidopsis thaliana (PSSU)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2658)..(3031))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4876))
; OTHER INFORMATION: PTA29: promoter region of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((4922)..(4946))
; OTHER INFORMATION: LB: T-DNA left border
US-08-817-188-1

```

```

Query Match      45.4%; Score 188.4; DB 3; Length 4946;
Best Local Similarity 99.5%; Pred. No. 1.4e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgccataagaagcaattgttagatgtaattcccatcttgaagaataat 285
Db 54 gtacatggtcgataagaagcaattgttagatgtaattcccatcttgaagaataat 113

```

```

QY 286 agtttaaatattattgataaaatacaagtcaggtattatagtcacagcaaaacataa 345
Db 114 agtttaaatattattgataaaatacaagtcaggtattatagtcacagcaaaacataa 173
QY 346 atttattgatcaagtttaaatcagaataattttcaataaactgatttatcagctggtag 405
Db 174 atttattgatcaagtttaaatcagaataattttcaataaactgatttatcagctggtag 233
QY 406 attgcccgtag 415
Db 234 attgcccgtag 243

RESULT 9
US-08-817-188-5
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopline TL-DNA from
; OTHER INFORMATION: pRI86S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' q7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: pSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-Ls1 gene
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopline TL-DNA from
; OTHER INFORMATION: pRI86S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; OTHER INFORMATION: nucleotides)
US-08-817-188-5

Query Match 45.4%; Score 188.4; DB 3; Length 5560;
Best Local Similarity 99.5%; Pred. No. 1.4e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 226 gtacatgcccgcataagaaggaatttgcagatgtgtaattcccatcttgaagaagaatat 285
Db 41 gtacatggtcgataagaaggaatttgcagatgtgtaattcccatcttgaagaagaatat 100
QY 286 agtttaaatattattgataaaatacaagtcaggtattatagtcacagcaaaacataa 345
Db 101 agtttaaatattattgataaaatacaagtcaggtattatagtcacagcaaaacataa 160
QY 346 atttattgatcaagtttaaatcagaataattttcaataaactgatttatcagctggtag 405
Db 161 atttattgatcaagtttaaatcagaataattttcaataaactgatttatcagctggtag 220
QY 406 attgcccgtag 415
Db 221 attgcccgtag 230

RESULT 10
US-08-232-016-23/c
; Sequence 23, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P. O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016

```

```
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pPS0212
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated modified bt2 (cryIAb) gene, also designated as the
; OTHER INFORMATION: cryIAb6 gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1793..2026
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2396..2921
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2922..3581
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3582..4407
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5600..6457
; OTHER INFORMATION: /note= "Sequence complementary to
; OTHER INFORMATION: Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7071..7566
; OTHER INFORMATION: /note= "TR1' and TR2 promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader with
; OTHER INFORMATION: respect to sequence of pJD884 of SEQ ID NO. 22."
; US-08-232-016-23
;
; Query Match 45.4%; Score 188.4; DB 2; Length 7566;
; Best Local Similarity 99.5%; Pred. No. 1.5e-36;
; Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 226 gtacatgcccgaataaaggcaattttagatgttaattcccatcttgaagaataat 285
; ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db 2070 GTACATGGTCGATAGAGAAAAGCAATTTCTAGATGTTAATTCCCATCTTGAAGAAATAT 2011
Qy 286 agttaaataattattgataaaatacaagtcaggtattattatgccaagcaaaaaacataa 345
; ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db 2010 ACTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGAAAACATAA 1951
Qy 346 atttattgatgcaagttttaattcaagaaatatttcaataactgattatatcagctggtac 405
; ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1950 ATTTATTGATCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATATCAGCTGTAC 1891
Qy 406 attgcgctag 415
; ||||| |||||||
Db 1890 ATTGCCGTAG 1881
;
; RESULT 11
; US-08-232-016-22/c
; Sequence 22, Application US/082332016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91402920.2
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pJD884
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1869
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated bt2 (cryIAb) gene, also designated as the bt884
; FEATURE:
; NAME/KEY: misc_feature
```

```
LOCATION: 1877..2110
OTHER INFORMATION: /note= '3' regulatory sequence
OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium
OTHER INFORMATION: T-DNA gene 7."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2480..3005
OTHER INFORMATION: /note= '35S promoter sequence
OTHER INFORMATION: derived from Cauliflower mosaic virus."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3006..3665
OTHER INFORMATION: /note= 'Coding sequence of
OTHER INFORMATION: chloramphenicol acetyl transferase gene."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3666..4491
OTHER INFORMATION: /note= '3' regulatory sequence
OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium
OTHER INFORMATION: T-DNA octopine synthase gene."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5684..6541
OTHER INFORMATION: /note= 'Sequence complementary to
Patent No. 5952547
OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7155..7639
OTHER INFORMATION: /note= 'TR1' and TR2' promoter
OTHER INFORMATION: derived from Agrobacterium T-DNA."
US-08-232-016-22
```

```
Query Match 45.4%; Score 188.4; DB 2; Length 7639;
Best Local Similarity 99.5%; Pred. No. 1.5e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 226 gtacatgcccatgaagaagaaggcaattgttagatgtaattcccatcttgaagaataat 285
Db 2154 GTACATGTCGATGAAGAAAGGCAATTTGTAGATGTTAATTCCTTGAAGAAATAT 2095
QY 286 agtttaaatattattgataaaatacaagtcagggtattatagtcgaagcaaaaacataa 345
Db 2094 AGTTTAAATATTATTGATAAAATAACAGTCAGGTATTATAGTCCCAAGCAAAACATAA 2035
QY 346 atttatgatcgaagtttaattcagaataatttcgaataactgattatcagctgtgac 405
Db 2034 ATTATTGATGCAAGTTTAAATTTCAGAAATATTTCATAACTGATTATATCATCGTGTAC 1975
QY 406 attgcccgtag 415
Db 1974 ATTGCCGTAG 1965
```

```
RESULT 12
US-08-894-440-4
Sequence 4, Application US/08894440
Patent No. 6025546
GENERAL INFORMATION:
APPLICANT: PLANT GENETIC SYSTEMS N.V.
TITLE OF INVENTION: Method to obtain male sterile plants
FILE REFERENCE: NMSCOR
CURRENT APPLICATION NUMBER: US/08/894,440
CURRENT FILING DATE: 1997-11-12
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 5864
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
```

```
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
```

```
OTHER INFORMATION: plasmid pTCO113
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement(11)..(25))
OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((98)..(330))
OTHER INFORMATION: region containing polyadenylation signal of gene 7
OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((331)..(882))
OTHER INFORMATION: region coding for phosphinothricin acetyl
OTHER INFORMATION: transferase (bar)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((883)..(2608))
OTHER INFORMATION: promoter of small subunit gene of Rubisco of
OTHER INFORMATION: Arabidopsis (Pssu)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((2659)..(3031))
OTHER INFORMATION: region containing polyadenylation signal of
OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((3032)..(3367))
OTHER INFORMATION: region coding for barnase of Bacillus
OTHER INFORMATION: amyloliquefaciens
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((3368)..(4877))
OTHER INFORMATION: promoter of nopaline synthase gene of
OTHER INFORMATION: Agrobacterium T-DNA (Phos)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5217)..(5489)
OTHER INFORMATION: region coding for barstar of Bacillus
OTHER INFORMATION: amyloliquefaciens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5490)..(5765)
OTHER INFORMATION: region containing polyadenylation signal of gene 7
OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((5840)..(5864))
OTHER INFORMATION: left border of Agrobacterium T-DNA
US-08-894-440-4
```

```
Query Match 45.0%; Score 186.8; DB 3; Length 5864;
Best Local Similarity 98.9%; Pred. No. 3.4e-36;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 226 gtacatgcccatgaagaagaaggcaattgttagatgtaattcccatcttgaagaataat 285
Db 54 gtacatggtcgataagaagaaggcaattgttagatgtaattcccatcttgaagaataat 113
QY 286 agtttaaatattattgataaaatacaagtcagggtattatagtcgaagcaaaaacataa 345
Db 114 agtttaaatattattgataaaatacaagtcagggtattatagtcgaagcaaaaacataa 173
QY 346 atttatgatcgaagtttaattcagaataatttcgaataactgattatcagctgtgac 405
Db 174 atttatgatcgaagtttaattcagaataatttcgaataactgattatcagctgtgac 233
```

Qy 406 attgccgtag 415  
|  
Db 234 attgccgtag 243

RESULT 13  
US-08-453-104-22/c  
; Sequence 22, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 42.7%; Score 177; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 239 aagaaaaaggcaattgttagatgtttaattcccatcttgaagaaataatagtttaaatattt 298  
|  
Db 3201 AAGAAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAAATATAGTTTAATATT 3142

Qy 299 attgataaaatacaagtcagggtattatagtcacagcaaaacataaattatttgatgca 358  
|  
Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAACATAAAATTTATTGATGCA 3082

Qy 359 agtttaaatcagaataatttcaataactgattatatacagctgggtacattgccgtag 415  
|  
Db 3081 AGTTTAATTCAGAAATATTTCATAACTGATTATATACAGCTGGTACATCCCGTAG 3025

RESULT 14  
US-08-694-824-22/c  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 42.7%; Score 177; DB 2; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 239 aagaaaaaggcaattgttagatgtttaattcccatcttgaagaaataatagtttaaatattt 298  
|  
Db 3201 AAGAAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAAATATAGTTTAATATT 3142

Qy 299 attgataaaatacaagtcagggtattatagtcacagcaaaacataaattatttgatgca 358  
|  
Db 3141 ATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAACATAAAATTTATTGATGCA 3082

```
Qy 359 agtttaattcagaataatttcaataactgattatatcagctggtacattgcoqtag 415
|||||
Db 3081 AGTTTAAATTCAGAAATATTTCATACTGATTATATCAGCTGGTACATTGCCGTAG 3025

RESULT 15
US-08-064-121-2/c
; Sequence 2, Application US/08064121
; Patent No. 5641664
; GENERAL INFORMATION:
; APPLICANT: D'HALLUIN, Kathleen
; APPLICANT: GOBEL, Elke
; TITLE OF INVENTION: PROCESS FOR TRANSFORMING
; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/064,121
; FILING DATE: 24-MAY-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90403332.1
; FILING DATE: 23-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91401888.2
; FILING DATE: 08-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: probe
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..8
; OTHER INFORMATION: /note= "sequence derived from
; OTHER INFORMATION: tapetum specific promoter of Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 9..790
; OTHER INFORMATION: /label= NPTII
; OTHER INFORMATION: /note= "coding sequence of neomycine
; OTHER INFORMATION: phosphotransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 791..1186
; OTHER INFORMATION: /label= 3'g7
; OTHER INFORMATION: /note= "3, regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
```

```
; OTHER INFORMATION: T-DNA gene 7"
US-08-064-121-2
```

```
Query Match 36.9%; Score 153; DB 1; Length 1186;
Best Local Similarity 100.0%; Pred. No. 2.8e-28;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 aattcccatcttgaagaataatagtttaaatattttgataaaataacaaagtcaggta 322
|||||
Db 1185 AATTCCCATCTTGAAAGAAATATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTA 1126

Qy 323 ttatagtcgaagcaaaacataaattttattgattgcaagtttaattcagaataatttcaa 382
|||||
Db 1125 TTATAGTCCAAGCAAAACATAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAA 1066

Qy 383 taactgatttatatcagctggtacattgcoqtag 415
|||||
Db 1065 TAACTGATTATATCAGCTGGTACATTGCCGTAG 1033
```

Search completed: February 25, 2002, 18:05:27  
Job time: 18575 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 17:21:04 ; Search time 8261.74 Seconds  
(without alignments)  
539.777 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagttgtgttcacga.....cagctgtacattgcgctag 415

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_estfun:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estom:*
5:	em_estpl:*
6:	em_estba:*
7:	em_estro:*
8:	em_estov:*
9:	em_htc:*
10:	gb_est1:*
11:	gb_est2:*
12:	gb_htc:*
13:	gb_gss:*
14:	em_gss_fun:*
15:	em_gss_hum:*
16:	em_gss_inv:*
17:	em_gss_pln:*
18:	em_gss_pro:*
19:	em_gss_rod:*
20:	em_gss_vrt:*
21:	em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	63	15.2	734 13 CNS010MP
2	61	14.7	1101 13 CNS0042W
3	57.2	13.8	844 13 CNS03D01
4	57.2	13.8	1101 13 CNS00088
5	57.2	13.8	1101 13 CNS00238
6	55.4	13.3	1101 13 CNS016LI
7	55	13.3	537 13 AQ506817
8	54.8	13.2	987 13 CNS014FQ
9	54.8	13.2	1101 13 CNS003BD
10	54.6	13.2	980 13 CNS00JG1
11	54.6	13.2	1101 13 CNS00EVL
12	54.4	13.1	893 13 CNS013XE

13	54	13.0	905	13	CNS00KHX	AL077798 Drosophil
14	53.4	12.9	500	10	AU087444	AU087444 Drosophil
15	53.4	12.9	970	13	CNS0182A	AL108796 Drosophil
16	53.2	12.8	1001	13	CNS0155H	AL105023 Drosophil
17	53.2	12.8	1101	13	CNS0106X	AL098595 Drosophil
18	53	12.8	1101	13	CNS016LI	AL106896 Drosophil
19	52.6	12.7	734	13	CNS010MP	AL099163 Drosophil
20	52.6	12.7	1013	13	CNS00J71	AL075824 Drosophil
21	52.6	12.7	1101	13	CNS00EPO	AL069493 Drosophil
22	52.4	12.6	524	13	CNS01U90	AL167541 Tetraodon
23	52.4	12.6	996	13	CNS00FUH	AL071063 Drosophil
24	52.2	12.6	639	13	CNS038CX	AL232458 Tetraodon
25	51.8	12.5	928	13	CNS00DKY	AL071865 Drosophil
26	51.8	12.5	1101	13	CNS0039G	AL063921 Drosophil
27	51.8	12.5	1101	13	CNS000FC	AL070972 Drosophil
28	51.6	12.4	992	13	CNS0562R	AL322812 Tetraodon
29	51.6	12.4	1043	13	CNS0145P	AL103735 Drosophil
30	51.6	12.4	1101	13	CNS00EO7	AL069440 Drosophil
31	51.4	12.4	876	13	CNS009G1	AL053529 Drosophil
32	51.2	12.3	678	13	CNS02A0C	AL187941 Tetraodon
33	51.2	12.3	1001	13	CNS01400	AL103554 Drosophil
34	51.2	12.3	1101	13	CNS00EVL	AL069706 Drosophil
35	50.6	12.2	854	11	BF274512	BF274512 GA_EB002
36	50.6	12.2	963	10	AL566565	AL566565 AL566565
37	50.6	12.2	1101	13	CNS00CVH	AL060100 Drosophil
38	50.6	12.2	1101	13	CNS0022U	AL097152 Drosophil
39	50.4	12.1	728	13	AQ272964	AQ272964 nbxb0028P
40	50.4	12.1	1101	13	CNS003BB	AL064089 Drosophil
41	50.4	12.1	1200	13	CNS016CO	AL106578 Drosophil
42	50.2	12.1	681	13	B60190	B60190 CIT-HSP-200
43	50	12.0	563	13	AQ326762	AQ326762 nbxb0038D
44	50	12.0	788	13	BH126604	BH126604 BARC-Satt
45	50	12.0	987	13	CNS014PQ	AL104456 Drosophil

ALIGNMENTS

RESULT 1

CNS010MP	734 bp	DNA	GSS	26-JUL-1999
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC			
DEFINITION	BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL099163			
VERSION	AL099163.1 GI:5610774			
KEYWORDS	GSS,			
SOURCE	fruit fly.			
ORGANISM	Plasmid Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 734)			
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)			
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (BDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.			
FEATURES	Location/Qualifiers			
	1..734			
	/organism="Drosophila melanogaster"			
	/plasmid="pBelOBAC11"			
	/db_xref="taxon:7227"			
	/clone_lib="DrosBAC"			
	/clone="BACN04L20"			

```

BASE COUNT      288 a      62 c      2 g      211 t      171 others
ORIGIN

Query Match
Best Local Similarity 15.2%; Score 63; DB 13; Length 734;
Matches 85; Conservative 43; Mismatches 79; Indels 0; Gaps 0;

QY 189 aagttattataattataattataattatggtaggatgtacatgccgataagaagaaggc 248
DB 53 AMATATTTATATATATATAATAATAATAAAMAMMTTCAATTTMHAATAAAWRAAW 112
QY 249 aattgttagtgaattcccatcttgaagaataatagttttaattattattgataaaa 308
DB 113 WTAWHAAATATTTWATTTTTHAAAWATATTTTMMATTTAAAAATTTTTTTAA 172
QY 309 taacagtcaggtattatagtcacagcaaacataaaatttattgatgcaagtttaaat 368
DB 173 AWMNMTYAWATTTTWTWATWAAAMWTWAAAWAAAWAAWATATATTATTTAA 232
QY 369 cagaataattccaataactgattatat 395
DB 233 AWMAATATWMAAAAAAMWWMMWTAT 259

RESULT 2
CNS004ZW/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC #
BAC11E08 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION AL055440
KEYWORDS AL055440.1 GI:4932241
SOURCE GSS.
ORGANISM
fruit fly
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrf@genosco.cns.fr
- web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/TheBDGP/Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Piter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BAC11E08"
/note="end : T7"

BASE COUNT      294 a      74 c      99 g      381 t      253 others
ORIGIN

Query Match
14.7%; Score 61; DB 13; Length 1101;

```

```
source
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR01A24"
/notes="end : TET3"

BASE COUNT      228 a  114 c  110 g  512 t  137 others
ORIGIN

Query Match      13.8%; Score 57.2; DB 13; Length 844;
Best Local Similarity 39.5%; Pred. No. 0.099;
Matches 155; Conservative 6; Mismatches 169; Indels 0; Gaps 0;

Qy 56 actcttaagtgagaaacactcaacagcattatcatctgttcataataatattgacat 115
Db 487 AGTGTTCGCCATAATCAAGTAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 428
Qy 116 tatcgttatatacacgtatcacaaatagtagcgaagaacccatgtaagcagcagg 175
Db 427 AATATAAAAAAATAAAATAAAATAATATTAATAAAATAAAATAATTAATAAAATAA 368
Qy 176 gcaccatggttcgaagtattataataattataattattggttaggtgacatggcc 235
Db 367 AAAAAAATAATATAAAAAATTAWAACACAAAAAATAAAATAAAACATAAAAAA 308
Qy 236 gataagaagaagcaatttgtagtggttaattcccatcttggaagaataatagtttaa 295
Db 307 AAAAAAATAAAATAAAATAATMAAAACMTAAATGATATAAAAWAAAAAATAAA 248
Qy 296 tttattgataaataacaagtcaggtattatagttcccaagcaaaacataaattttgat 355
Db 247 AAAAAAATAAAATAATACAMATAAAATAAAATAAAATAAAATAAAATAAAATAA 188
Qy 356 gcaagtttaattcagaataattccaataa 385
Db 187 TCTAAAAAATCATAAACATTAAATCA 158

RESULT 4
CNS000B8/c 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR01A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063632
VERSION AL063632.1 GI:4938680
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
Genoscope.
Direct Submission
TITLE Drosophila melanogaster genome survey sequence
AUTHORS BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. .1101
```

```
source
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR01A24"
/notes="end : TET3"

BASE COUNT      228 a  114 c  110 g  512 t  137 others
ORIGIN

Query Match      13.8%; Score 57.2; DB 13; Length 1101;
Best Local Similarity 39.5%; Pred. No. 0.094;
Matches 147; Conservative 56; Mismatches 165; Indels 4; Gaps 2;

Qy 43 attacattgaaactcttacggtagagaaacacacacagcattatcatgttcata 102
Db 1087 AWAAWAAATTAATTTTWWAAAAAATAAAATAAAATAAAATAAAATAAAATAA 1029
Qy 103 aatatactacattatcgttatatacacgtatatacaatagtagcgaagaataccatgt 162
Db 1028 AAAAAATAAATTTTAATATAAAATAAAATAAAATAAAATAAAATAAAATAAA 969
Qy 163 aaagcagcaggggccaccatggtttcaagtattataataattataattatggtag 222
Db 968 ATWTAWAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 909
Qy 223 gatgtacatggcgataagaagaaggcaattttagatgttaattcccatcttgaagaaa 282
Db 908 TTTAATWATWAAAAATWAAAAATTTTAATWTTAAATTTTAAATTTTAAATTA 849
Qy 283 tatagtttaaat---attattgataaaatacaacagtcaggtattatgccaaagcaaa 339
Db 848 AATWAAAAAATAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAA 789
Qy 340 acataaatttattgatgcaagtttaaatccagaataatttccataactatataatcagc 399
Db 788 ATWAAAAATACMAAAATYAAATYAAAAATAAAATAAAATAAAATAAAATAAA 729
Qy 400 tgggtacattgcc 411
Db 728 AATTAWATTHCM 717

RESULT 5
CNS00238/c 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN01A24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL097166
VERSION AL097166.1 GI:5608777
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
Genoscope.
Direct Submission
TITLE Drosophila melanogaster genome survey sequence
AUTHORS BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
Location/Qualifiers
1. .1101
```

```
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db_xref="taxon:7227"  
/clone_lib="DrosBAC"  
/clone="BACN01A24"  
/note="end : sp6"
```

BASE COUNT	308 a	152 c	162 g	386 t	93 others
ORIGIN					

Query Match 13.8%; Score 57.2; DB 13; Length 1101;  
Best Local Similarity 41.4%; Pred. No. 0.094;  
Matches 87; Conservative 38; Mismatches 85; Indels 0; Gaps 0;

[illegible][illegible]

Qy	306	aaataa	caagtc	caggtatt	tattatg	tcagcg	caaaaac	ataaaat	ttattat	tgatgc	aaagtt	taa	365
		:: ::	:: ::	:: ::	:: ::	:: ::	:: ::	:: ::	:: ::	:: ::	:: ::	:: ::	
Db	918	WTTTAAAA	AAAAAAAA	AAAAAAAA	AAAAAAAA	AAAAAAAA	AAAAAAAA	AAAAAT	MAATTA	TAAATTA	TTAAATTA	TTAAATTA	859

Qy 366 attcagaataattttcaataactgattatat 395  
          ||| : ||| : ::::: | : |  
Db 858 WTWTWTWTWTWTMMMHMHHAAYTAAWT 829

[illegible]

ACCESSION AL106896  
VERSION AL106896.1 GI:5624374  
KEYWORDS GSS.  
SOURCE fruit fly.

ORGANISM  
Eukaryota; Metazoa; Artropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)

REFERENCE

Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqsfr@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelOBAC11.

```

FEATURES
  source
    Location/Qualifiers
      1..1101
      organism="Drosophila
      plasmid="pBeloBac11
      gb_xref="taxon:7227"
      clone_lib="DrosBAC"
      clone="BACN16D22"
      notes="end..87"

```

BASE COUNT	203 a	220 c	84 g	158 t	436 others
RIGIN					

Query Match 13.3%; Score 55.4; DB 13; Length 1101;  
Best Local Similarity 30.7%; Pred. No. 0.21;

Matches	96;	Conservative	79;	Mismatches	138;	Indels	0;	Gaps	0;
Qy	43	attacattataaaactcttacgtagtagagaacaactcacagcattataatcgtgtccatata	102						
		:    :    :    :    :    :    :    :    :    :    :    :    :    :							
Db	1050	AWKAAWAAWAAWAAWTTTAYATTHTWTAATHTAWTWYCAATAHWWAWTTTATATWAAHT	991						
Qy	103	aatatgtatcatcgttatatacacgqtatacaaatagtagcgagaatacccatgt	162						
		:    :    :    :    :    :    :    :    :    :    :    :    :							
Db	990	TATTWATAHTATYTWTHWATYWWTWYTWATWAWAHAAWAAWTAATAAYWHATAATW	931						
Qy	163	aaagcagcagggggcaacatggtttcaagtattataataattataataattatggtag	222						
		:    :    :    :    :    :    :    :    :    :    :    :    :							
Db	930	AWAWATAATTTWTTWTTAWTTWAWTAATATWTAATWAAWTAATWAAWAWHWATAYAWW	871						
Qy	223	gatgtacatggccatagaagaaagccaatttggtagtgcattcccatcttgcagaaa	282						
		:    :    :    :    :    :    :    :    :    :    :    :    :							
Db	870	WWNAWTATWTTATWATAATAWWATTTWTAATATATWTAATAATATATATWTAATAAAWAA	811						
Qy	283	tatagtttaaatattttatgataaaaatacaagtcagggtattatagtcacagcaaaaaca	342						
		:    :    :    :    :    :    :    :    :    :    :    :    :							
Db	810	THAAWAAWAAATATWAAAAAAAHHTTAAAWATAAAAAWAAAAWAAAAWAAAAH	751						
Qy	343	taaatcttcatgtat	355						
		:    :    :    :    :    :    :    :    :    :    :    :    :							
Db	750	ATMAATTTTTTTTWT	738						

RESULT	7			
AQ506817/c				
LOCUS	AQ506817	537 bp	DNA	GSS
DEFINITION	RPCI-11-28J17.TV RPCI-11 Homo sapiens genomic clone RPCI-11-28J17			
	, DNA sequence.			

ACCESSION AQ506817  
VERSION AQ506817.1  
KEYWORDS GSS.  
SOURCE human.

ORGANISM      Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE      1 (bases 1 to 537)

**AUTHORS** Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

**TITLE** Use of PAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

**JOURNAL** unpublished (1997)

Other\_GSSs: RPCI-11-281j17\_TJ  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research

9/12 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org

Clones are derived from the human BAC library RPOI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page: [http://www.tigr.org/tldb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html). Seq primer: 77

```

Class: BAC ends.
FEATURES
      Location/Qualifiers
      1..537
      source

```

```

/organism="Homo sapiens"
/db_xref="GDB:7607732"
/db_xref="taxon:9606"
/clone="RPC1-11-281J17"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/notice_vector="phages3.6; Site_1: ECORI;
RPC111 Human Male BAC Library"
216 a 30 q 280 t
BASE COUNT

```

## ORIGIN

Query Match 13.3%; Score 55; DB 13; Length 537;  
Best Local Similarity 51.0%;  
Matches 181; Conservative 0; Mismatches 170. Indels  
Pred. No. 0.28;

Qy	45	tacatttgaaactcttaaggatgagaacaactcacaaagcatttaactctgttcataaaa	104
Db	453	TATATATAAATATATATATAAATATATATATATATATAAATATATATAA	394
Qy	105	tatatgtacat - tatacgtatatacacogtatacacaaatagtagcgagaagaatccatgta	163
Db	393	TATATACATATATATATATATAAATATATAAATATATATATATAAATATACATATA	334
Qy	164	aagcagcagggggccacat--gggttccaagtattataaattataattataattatggt	220
Db	333	TATCCATATATATAAATATATACATATATATAAATATATATAAATATATAATATA	274
Qy	221	aggatgtacatggccgataaagaaaggcaattgttagtgttaattcccattcttgaaaga	280
Db	273	AATATATATACATATATCAATATATATCAATATATATAAATATATATCAATATATAAATA	214
Qy	281	aetatagttcaaatatttatttgataaaataaacaagtcagggttattagtcacagcaaaaa	340
Db	213	TATATAAATATATATAAATATATATAAATATATAAATATATATACATATATATAAATATA	154
Qy	341	cataaaatttattgatgcaaggtttaaattcagaagaattttcaataacgattatat	395
Db	153	TATAAATATATAAATATATATAAATATAAATATATAAATATATATATATATACATATAT	99

**RESULT** 8

CNS014PQ	987 bp	DNA	GSS	26-JUL-1999
LOCUS				
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL104456			
VERSION	AL104456			
KEYWORDS	AL104456.1	GI:5616067		
SOURCE	GSS.			
ORGANISM	fruit fly. Plasmid Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 987)			
REFERENCE				

COMMENT

web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
determination of this BAC-end sequence was carried out as part of a  
collaboration with the Human Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelOBC11.

Best Local Similarity 26.8%; Pred. No. 0.27;  
Matches 103; Conservative 100; Mismatches 182; Indels 0; Gaps 0;

[illegible]

## RESULT 9

CNS003BD	CNS003BD	1101 bp	DNA	GSS	03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #				
DEFINITION	BACR08K08 of RFI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL064091				
VERSION	AL064091.1				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				

[illegible]



[illegible]



## ORIGIN

[illegible]

RESULT 14  
 LOCUS AU087444/c  
 DEFINITION  
 AU087444 500 bp mRNA  
 AU087444 Sugano Malaria cDNA library  
 clone XPrn4669, mRNA sequence.  
 EST  
 AU087444  
 AU087444.1 GI:12389585  
 EST  
 malaria parasite P. falciparum.  
 Plasmodium falciparum  
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 1 (bases 1 to 500)  
 Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.  
 FULL-malaria: a database for a full-length enriched cDNA library  
 from human malaria parasite, plasmodium falciparum  
 Nucleic Acids Res. 29 (1), 70-71 (2001)  
 20574754  
 Contact: Junichi Watanabe

Email: jwatanabe@ange.ims.u-tokyo.ac.jp  
Suzuki.Y., Yoshitomo-Nakagawa.K., Maruyama.K., Suyama.A. and Sugano.S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES	source
a 3'-end-enriched CDNA library Gene 200 (1-2), 149	
Location/Qualifiers	
1..500	
/organism="Plasmodium falciparum"	
/strain="3D7"	
/db_xref="taxon:5833"	
/clone_lib="XPFn4669"	
/dev_stage="Sugano Malaria CDNA library"	
203 a 36 c 24 g 227 t 10 others	
BASE COUNT	
gcg	

Query Match 12.9%; Score 53.4; DB 10; Length 500;  
Best Local Similarity 50.3%; Pred. NO. 0.57;  
Matches 155; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

89 atcatggttcataaaatataatgtacattatacgtatatatatacacggtatacaaaatagtag 147

Db	499	ATAATTTAAATATAAAATAAATAACATACATAGATATATATATATATATATTTATATCGT	440
Qy	148	cgaagaataccatgtaaaagcagcagggggccaccatggtttcaggtattataaattataa	207
Db	439	TTAATAATACAGTATATTTTATACGTATATGTCGCCCTCAATATTTTATTAATATAT	380
Qy	208	ttataaattatggtaggatgtcacatggccgataagaagaaaggccaatttgcagtgttaattc	267
Db	379	TTAGAAAAATAGCTTTATCTTAAAAATATAAAGAGAAAATTTTATTTTATATATTATACN	320
Qy	268	ccatcttgaagaataatagttttaaataatttattcgtataaaatacaacagtcaggattata	327
Db	319	AAVATAITTTTATATCTATTAATATATATATACATGCATAGGAAATAAAAATAAAA	260
Qy	328	gtccagcaaaaacataaatttattgatgcaggtttaaattcagaataatttcgaataact	387
Db	259	ATATACATATATATATATATATATATATATATATATATAGTAAAAAGGATTTGGCTATTATA	200
Qy	388	gattatat	395
Db	199	AAATATTAT	192

RESULT	15
CNS0182A	GSS
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC BAC37C16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
DEFINITION	970 bp DNA

ACCESSION	AL108796
VERSION	AL108796.1
KEYWORDS	GI:5629100
SOURCE	GSS.
ORGANISM	fruit fly. Plasmod Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 970)

AUTHORS	TITLE	JOURNAL	COMMENT
Genoscope. Direct Submission	Submitted (23-JUL-1999) BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)	Genoscope - Centre National de Sequencage : Genome project (EDGP) - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billault at CSHG (Centre d'etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.	

FEATURES	source	Location/Qualifiers
peloBAC11		1..970
		/organism="Drosophila melanogaster"
		/plasmid="peloBAC11"
		/db_xref="taxon:7227"
		/clone_lib="DrosBAC"
		/clone="BACN37C16"
		/note="end c T7"
BASE COUNT	458 a	107 c 81 g 124 t 200 others
ORIGIN		

Query Match 12.9%; Score 53.4; DB 13; Length 970;  
Best Local Similarity 42.6%; Pred. No. 0.51;  
Matches 144; Conservative 12.9%;

	Machine	144;	Conservative	13;	Mismatches	181;	Indels	0;	Gaps	0;
39	2y	caccattacattgaaactcttcacggatgagaacaactcacaagcattatcatgttca	98							
15	0b	CAAAATTTNNATNNNNTTAAATTTNNNAAAAAAAAAAAAAAAAAAATAATTTAA	74							
99	2y	tataataatgacattacgttatatacacgttatatacaaatagtcagcaagaatcc	158							



```
Db 75 AAAAAAAAAACCCCTAATAAATCNACAAAAATTTAAAAATAATAATAAAAAATAAAA 134
Qy 159 atgtaagcagcagggggcaccatggtttcaagtattatataattataattatg 218
   ||:|||||:| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 ATAWAAAAAAATWTTTATWAAWAAAAATATAWATAWTAATAAAAAAANTTNCTAATTTT 194
Qy 219 gtaggatgtacatggcgcatagaagaaaggcaattttagatgttaattcccatcttgaaa 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 ATTATNATAAATAANCTATAAAAAAATAAAAAATAAAAAATAAAAAACAATATAAAAA 254
Qy 279 gaaatatagtttaaatattttatgataaaaataacaagtcagggtattatagtcacaagcaaa 338
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 ATAAAAAATAAATAAAAAAATAAATAAAAAAATAAAAAAATAAAAAAATAAAAAATAAATA 314
Qy 339 aacataaatttattgatgcaagtttaaatcagaata 376
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 AAAAATAAATAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 352
```

Search completed: February 25, 2002, 17:21:06  
Job time: 16159 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:03:22 ; Search time 1457.19 Seconds  
(without alignments)  
4698.309 Million cell updates/sec

Title: US-09-698-903B-8

Perfect score: 415

Sequence: 1 gtcgagttgtgttcattga.....cagctgtacattgcgctag 415

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_cm.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\*

Result No.	Score	Query Match	Length	DB ID	Description
1	415	100.0	415	6	AX127755 Sequence
2	380.8	91.8	415	6	AX127757 Sequence
3	191.4	46.1	5865	6	AX127748 Sequence
4	191.4	46.1	7599	6	AX063413 Sequence
5	188.8	45.5	6539	6	E31991 Mutated bar
6	188.8	45.5	6548	6	A60109 Sequence 2
7	188.8	45.5	6548	6	A76916 Sequence 2
8	188.8	45.5	6548	6	AR098308 Sequence
9	188.8	45.5	6548	6	E31990 Mutated bar
10	188.8	45.5	7811	6	AR078675 Sequence
11	188.4	45.4	3200	6	I44104 Sequence 23
12	188.4	45.4	4832	6	AX172441 Sequence
13	188.4	45.4	4946	6	A60108 Sequence 1
14	188.4	45.4	4946	6	A76915 Sequence 1
15	188.4	45.4	4946	6	AR098307 Sequence
16	188.4	45.4	4946	6	AX172440 Sequence 7
17	188.4	45.4	5349	6	A71437 Sequence 7
18	188.4	45.4	5560	6	A60112 Sequence 5
19	188.4	45.4	5560	6	AR098311 Sequence
20	188.4	45.4	5865	6	AX127748 Sequence
21	188.4	45.4	7566	6	A24783 Sequence
22	188.4	45.4	7566	6	AR074388 Sequence
23	188.4	45.4	7639	6	A24782 plasmid pJD
24	188.4	45.4	7639	6	AR074387 Sequence
25	182.4	44.0	1037	6	A10942 Nucleotide
26	182.4	44.0	1085	6	A10939 Nucleotide
27	182.4	44.0	1160	6	A10943 Nucleotide
28	179.4	43.2	1077	6	AX172463 Sequence 22
29	177	42.7	3201	6	I44103 Sequence 22
30	155	37.3	2476	12	TBI251013 Transform
31	155	37.3	3236	12	TBI251014 Transform
32	153	36.9	1186	6	AI8051 DNA used as
33	153	36.9	1186	6	AR095107 Sequence
34	153	36.9	1186	6	AR098313 Sequence
35	153	36.9	1186	6	AX012338 Sequence 2
36	153	36.9	1186	6	I49886 Sequence 2
37	153	36.9	1186	6	I82374 Sequence 2
38	153	36.9	12095	12	BINHYGDNA
39	146	35.2	831	1	ATRN7
40	146	35.2	878	1	ATDNA
41	146	35.2	24595	1	ATACH5
42	146	35.2	24595	6	E00404 T1 plasmid
43	146	35.2	24595	6	E00546 DNA fragmen
44	146	35.2	194140	1	AF242881 Agrobacte
45	116	28.0	249	12	ARGMTUB X05579 Soybean bet

ALIGNMENTS

RESULT 1

AX127755

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

PAT 15-MAY-2001

AX127755 Sequence 8 from Patent WO0131042.

AX127755

AX127755.1 GI:14134402

synthetic construct.

artificial construct

artificial sequence.

1 (bases 1 to 415)

Weston,B. and de Beuckeleer,M.

Male-sterile brassica plants and methods for producing same

Patent: WO 0131042-A 8 03-MAY-2001;

Avantis CropScience N.V. (BE)

Location/Qualifiers

1..415

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="5' border flanking region of elite event MS-B2"

```
misc_feature 1..234
/note="plant DNA"
misc_feature 235..415
/note="r-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN

Query Match 100.0%; Score 415; DB 6; Length 415;
Best Local Similarity 100.0%; Pred. No. 2.2e-64;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgagtttggttcacgatttgggtttgactcttccacattacattgaactctt 60
DB 1 gtcgagtttggttcacgatttgggtttgactcttccacattacattgaactctt 60
QY 61 tacgagtgagaaactcacagcatttaattcattgctcattataataattgacattat 120
DB 61 tacgagtgagaaactcacagcatttaattcattgctcattataataattgacattat 120
QY 121 gatatatacgcgtatacaaatagtagcgagaataatccatgtaagcagcgagggcgacc 180
DB 121 gatatatacgcgtatacaaatagtagcgagaataatccatgtaagcagcgagggcgacc 180
QY 181 atggtttcaagttattataattataattatggttaggtgtaggtgacatgcccgtata 240
DB 181 atggtttcaagttattataattataattatggttaggtgtaggtgacatgcccgtata 240
QY 241 gaaagggcaatttgatgatttccatcttccatcttgaagaaataatgatttataattattat 300
DB 241 gaaagggcaatttgatgatttccatcttccatcttgaagaaataatgatttataattattat 300
QY 301 tgataataatacagtcaggtattattatagtcgaagcaaaacataataatttattgacgaag 360
DB 301 tgataataatacagtcaggtattattatagtcgaagcaaaacataataatttattgacgaag 360
QY 361 ttttaattcagaataatttcaataactgatttatcagctggtacattgcccgtag 415
DB 361 ttttaattcagaataatttcaataactgatttatcagctggtacattgcccgtag 415

RESULT 2
AX127757/c 416 bp DNA PAT 15-MAY-2001
LOCUS AX127757
DEFINITION Sequence 10 from Patent WO0131042.
ACCESSION AX127757
VERSION AX127757.1 GI:14134404
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 416)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 10 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
    Location/Qualifiers
    source
        1..416
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="3' border flanking region of elite event MS-B2"
misc_feature 1..193
/note="r-DNA"
misc_feature 194..416
/note="plant DNA"
BASE COUNT 137 a 72 c 54 g 152 t 1 others
ORIGIN

Query Match 91.8%; Score 380.8; DB 6; Length 416;
Best Local Similarity 98.5%; Pred. No. 2.5e-58;
Matches 404; Conservative 1; Mismatches 3; Indels 2; Gaps 2;

misc_feature 1..234
/note="plant DNA"
misc_feature 235..415
/note="r-DNA"
BASE COUNT 154 a 55 c 70 g 136 t
ORIGIN

Query Match 100.0%; Score 415; DB 6; Length 415;
Best Local Similarity 100.0%; Pred. No. 2.2e-64;
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgagtttggttcacgatttgggtttgactcttccacattacattgaactctt 60
DB 1 gtcgagtttggttcacgatttgggtttgactcttccacattacattgaactctt 60
QY 61 tacgagtgagaaactcacagcatttaattcattgctcattataataattgacattat 120
DB 61 tacgagtgagaaactcacagcatttaattcattgctcattataataattgacattat 120
QY 121 gatatatacgcgtatacaaatagtagcgagaataatccatgtaagcagcgagggcgacc 180
DB 121 gatatatacgcgtatacaaatagtagcgagaataatccatgtaagcagcgagggcgacc 180
QY 181 atggtttcaagttattataattataattatggttaggtgtaggtgacatgcccgtata 240
DB 181 atggtttcaagttattataattataattatggttaggtgtaggtgacatgcccgtata 240
QY 241 gaaagggcaatttgatgatttccatcttccatcttgaagaaataatgatttataattattat 300
DB 241 gaaagggcaatttgatgatttccatcttccatcttgaagaaataatgatttataattattat 300
QY 301 tgataataatacagtcaggtattattatagtcgaagcaaaacataataatttattgacgaag 360
DB 301 tgataataatacagtcaggtattattatagtcgaagcaaaacataataatttattgacgaag 360
QY 361 ttttaattcagaataatttcaataactgatttatcagctggtacattgcccgtag 415
DB 361 ttttaattcagaataatttcaataactgatttatcagctggtacattgcccgtag 415

RESULT 2
AX127748/c 5865 bp DNA PAT 15-MAY-2001
LOCUS AX127748
DEFINITION Sequence 1 from Patent WO0131042.
ACCESSION AX127748
VERSION AX127748.1 GI:14134395
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 5865)
AUTHORS Weston,B. and de Beuckeleer,M.
TITLE Male-sterile brassica plants and methods for producing same
JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;
Aventis CropScience N.V. (BE)
FEATURES
    Location/Qualifiers
    source
        1..5865
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="r-DNA of plasmid pCO113"
BASE COUNT 1849 a 1095 c 1149 g 1772 t
ORIGIN

Query Match 46.1%; Score 191.4; DB 6; Length 5865;
Best Local Similarity 99.5%; Pred. No. 5.7e-25;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 223 gatgtacatgcccgaataagaaagcaattttagatgttaattcccatcttgaagaaa 282
DB 5813 GATGTACATGTCGATAGACAAAGCAATTTGTAGATGTTAATCCCATCTGTAAGAAA 5754
QY 283 tatagtttaaatattatttgataaaatacaagtcaggtattattagtcacagcaaaaaca 342
DB 5753 TATAGTTTAAATATTATTGATTAATAAATAACAAGTCAGGTATTATAGTCCAGCAAAA 5694
QY 343 taaatttattgacgaagtttaaaattcagaataatttcaataactgattatcagctgg 402
DB 5693 TAAATTTATTGACGAAGTTTAAATTCAGAAATATTTCATTAACACTGATTATATCAGCTGG 5634
QY 403 tacattgcccgtag 415
DB 5633 TACATTGCCGTAG 5621
```

RESULT 4  
AX063413/c  
LOCUS AX063413 7599 bp DNA PAT 24-JAN-2001  
DEFINITION Sequence 5 from Patent WO0100833.  
ACCESSION AX063413  
VERSION AX063413.1 GI:12541201  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 7599)  
AUTHORS Hoffmann,B., Moillier,P. and Pelletier,G.  
TITLE Promoter expressed specifically in the cells of plant roots,  
recombinant vectors and host cells comprising same and transgenic  
plants obtained  
JOURNAL Patent: WO 0100833-A 5 04-JAN-2001;  
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (I.N.R.A.) (FR)  
FEATURES  
source Location/Qualifiers  
1. .7599  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="ADN-T de PGK85"  
BASE COUNT 1972 a 1938 c 1937 g 1752 t  
ORIGIN  
Query Match 46.1%; Score 191.4; DB 6; Length 7599;  
Best Local Similarity 99.5%; Pred. No. 5.6e-25;  
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 223 gatgtacatgcccgcataagaagggcaattttagatgtttaattcccatcttgaagaaa 282  
Db 6771 GATGTACATGTCGATAGAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAA 6712  
Qy 283 tatagtttaaatattttatgataaaatacaagtcaggattatttagtcccaagcaaaaaca 342  
Db 6711 TATAGTTTAAATATTTTATTGATAAAATTAACAAGTCAGGTATTATAGTCCAAGCAAAAACA 6652  
Qy 343 taaatttatgatgcagatttaattcagaaataatttcataactgatttatcatcagctgg 402  
Db 6651 TAAATTTATGTCAGCAAGTTTAAATTCAGAAATATTTCAATCACTGATTATATCAGCTGG 6592  
Qy 403 tacattgcccgtag 415  
Db 6591 TACATTGCCGTAG 6579  
RESULT 5  
E31991/c  
LOCUS E31991 6539 bp DNA PAT 07-FEB-2001  
DEFINITION Mutated barnase gene and transgenic plant thereof.  
ACCESSION E31991  
VERSION E31991.1 GI:13021588  
KEYWORDS JP 2000041682-A/4.  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 6539)  
AUTHORS Kazuyuki,H.F.N.N.  
TITLE Mutated barnase gene and transgenic plant thereof  
JOURNAL Patent: JP 2000041682-A 4 15-FEB-2000;  
JAPAN TOBACCO INC  
COMMENT OS Escherichia coli LE392  
PN JP 2000041682-A/4  
PD 15-FEB-2000  
PF 04-AUG-1998 JP 1998220060  
PR KAZUYUKI HAMADA,FUMIO NAKAKIDO  
PC C12N15/09,A01H5/00,C12N5/10,C12N9/22/(C12N5/10,C12R1:91), PC  
C12N15/00,  
PC C12N5/00, (C12N5/00,C12R1:91)  
CC

FEATURES  
source Location/Qualifiers  
1. .6539  
/organism="Escherichia coli LE392".  
BASE COUNT 1755 a 1578 c 1519 g 1687 t  
ORIGIN  
Query Match 45.5%; Score 188.8; DB 6; Length 6539;  
Best Local Similarity 99.0%; Pred. No. 1.6e-24;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 224 atgtacatgcccgcataagaaggaattttagatgtttaattcccatcttgaagaaaat 283  
Db 6533 ACGTACATGGTCGATAGAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAAAT 6474  
Qy 284 atagtttaaatattttatgataaaatacaagtcagggtatttatagtcacaagcaaaaacat 343  
Db 6473 ATAGTTTAAATATTTATTGATAAAATACAAAGTCAGGTATTATAGTCCAAGCAAAAACAT 6414  
Qy 344 aaatttatgatgcagtttaattcagaataatttcaataactgatttatcatcagctggt 403  
Db 6413 AAATTTATGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGGT 6354  
Qy 404 acattgcccgtag 415  
Db 6353 ACATTGCCGTAG 6342  
RESULT 6  
A60109/c  
LOCUS A60109 6548 bp DNA circular PAT 22-OCT-1999  
DEFINITION Sequence 2 from Patent WO9706267.  
ACCESSION A60109  
VERSION A60109.1 GI:3715125  
KEYWORDS  
SOURCE Plasmid PTS172.  
ORGANISM Plasmids.  
REFERENCE 1 (bases 1 to 6548)  
AUTHORS De,B.M.  
TITLE Genetic transformation using a PARP inhibitor  
JOURNAL Patent: WO 9706267-A 2 20-FEB-1997;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES  
source Location/Qualifiers  
1. .6548  
/organism="Plasmid PTS172"  
/db\_xref="taxon:106340"  
BASE COUNT 1756 a 1579 c 1523 g 1690 t  
ORIGIN  
Query Match 45.5%; Score 188.8; DB 6; Length 6548;  
Best Local Similarity 99.0%; Pred. No. 1.6e-24;  
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 224 atgtacatgcccgcataagaaggaattttagatgtttaattcccatcttgaagaaaat 283  
Db 6542 ACGTACATGGTCGATAGAAGGCAATTTGTAGATGTTAAATCCCATCTTGAAGAAAAT 6483  
Qy 284 atagtttaaatattttatgataaaatacaagtcagggtatttatagtcacaagcaaaaacat 343  
Db 6482 ATAGTTTAAATATTTATTGATAAAATACAAAGTCAGGTATTATAGTCCAAGCAAAAACAT 6423  
Qy 344 aaatttatgatgcagtttaattcagaataatttcaataactgatttatcatcagctggt 403  
Db 6422 AAATTTATGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGGT 6363  
Qy 404 acattgcccgtag 415  
Db 6363 ACATTGCCGTAG 6342

Db	6542	ACGTACATGGTCGATAGAAAAGGCAATTTGTAGTGAATTTCCATCTTCAAGAAT	6483
QY	284	atagtttaaatattattattgataaaatacaagtcaggtattatagtcgaagcaaaacac	343
Db	6482	ATAGTTTAAATATTTATTGATATAAATAACAAGTCAGGTATTAGTCCAAAGCAAAAACAT	6423
QY	344	aaatttattgatgcaagtttaaatcagaataattcctaactgattatcagctggt	403
Db	6422	AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGTT	6363
QY	404	acattgccgtag 415	
Db	6362	ACATTGCCGTAG 6351	
RESULT	9		
LOCUS	E31990/c	E31990 6548 bp DNA PAT 07-FEB-2001	
DEFINITION		Mutated barnase gene and transgenic plant thereof.	
ACCESSION	E31990		
VERSION	E31990.1	GI:13021587	
KEYWORDS	JP 2000041682-A/3.		
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 6548)		
AUTHORS	Kazuyuki H.F.N.N.		
TITLE	Mutated barnase gene and transgenic plant thereof		
JOURNAL	Patent: JP 2000041682-A 3 15-FEB-2000;		
COMMENT	JAPAN TOBACCO INC		
	OS Escherichia coli LE392		
	PN JP 2000041682-A/3		
	PD 15-FEB-2000		
	PF 04-AUG-1998 JP 1998220060		
	PR		
	PI KAZUYUKI HAMADA,FUMIO NAKAKITDO		
	PC C12N15/09,A01H5/00,C12N5/10,C12N9/22/((C12N5/10,C12R1:91),PC,C12N15/00,		
	CC C12N5/00,(C12N5/00,C12R1:91)		
FEATURES			
source			
FT	Key	Location/Qualifiers	
	FT source	1..6548	
		/organism="Escherichia coli LE392".	
BASE COUNT	1756 a 1579 c 1523 g 1690 t		
ORIGIN			
Query Match	45.5%;	Score 188.8;	DB 6; Length 6548;
Best Local Similarity	99.0%;	Pred. No. 1.6e-24;	
Matches 190;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
QY	224	atgtacatggcgcgaataagaagcgcaattttagatgttaattcccatcttgaagaacat	283
Db	6542	ACGTACATGGTCGATAAGCAAGGCAATTGTCAGATGTTAATTCCTATCTTGAAAGAAAT	6483
QY	284	atagtttaaatattattattgataaaatacaagtcaggtattatagtcgaagcaaaacac	343
Db	6482	ATAGTTTAAATATTTATTGATATAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAAACAT	6423
QY	344	aaatttattgatgcaagtttaaatcagaataattcctaactgattatcagctggt	403
Db	6422	AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGTT	6363
QY	404	acattgccgtag 415	
Db	6362	ACATTGCCGTAG 6351	



FEATURES  
source  
PLANT GENETIC SYSTEMS NV (BE)  
Location/Qualifiers  
1. .4946

/organism="Transformation vector pTHW107"  
/db\_xref="taxon:126810"  
BASE COUNT 1569 a 891 c 963 g 1523 t

ORIGIN

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaagaaggaattttagatgttattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGGTCGATAAGAAAGCAATTGTAGATGTTAAATCCCATCTTGAAAGAAATAT 113  
QY 286 agtttaaatatttattgataaaataacaagtcaggtattatagtcgaagcaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACCAAGTCAGGTATTATAGTCCAAAGCAAAACATAA 173  
QY 346 atttattgatgcgaatttcaattcagaataatttcaataactgattatcagctggtac 405  
|||||  
Db 174 ATTUATTCATCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGTAC 233  
QY 406 attgccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

RESULT 14  
LOCUS A76915 4946 bp DNA PAT 19-OCT-1999  
DEFINITION Sequence 1 from Patent EP0757102.  
ACCESSION A76915  
VERSION A76915.1 GI:6088712  
KEYWORDS Transformation vector pTHW107.  
SOURCE Transformation vector pTHW107.  
ORGANISM Transformation vector pTHW107.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De,B.M.  
TITLE GENETIC TRANSFORMATION USING A PARP INHIBITOR  
JOURNAL Patent: EP 0757102-A 1 05-FEB-1997;  
PLANT GENETIC SYSTEMS NV (BE)  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="Transformation vector pTHW107"  
/db\_xref="taxon:126810"  
BASE COUNT 1569 a 891 c 963 g 1523 t

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaagaaggaattttagatgttattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGGTCGATAAGAAAGCAATTGTAGATGTTAAATCCCATCTTGAAAGAAATAT 113  
QY 286 agtttaaatatttattgataaaataacaagtcaggtattatagtcgaagcaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACCAAGTCAGGTATTATAGTCCAAAGCAAAACATAA 173  
QY 346 atttattgatgcgaatttcaattcagaataatttcaataactgattatcagctggtac 405  
|||||  
Db 174 ATTUATTCATCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGTAC 233  
QY 406 attgccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

RESULT 15  
LOCUS AR098307  
DEFINITION Sequence 1 from patent US 6074876.  
ACCESSION AR098307  
VERSION AR098307.1 GI:12807564  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4946)  
AUTHORS De Block,M.  
TITLE Genetic transformation using a PARP inhibitor  
JOURNAL Patent: US 6074876-A 1 13-JUN-2000;  
FEATURES Location/Qualifiers  
source  
1. .4946  
/organism="unknown"  
BASE COUNT 1569 a 891 c 963 g 1523 t

Query Match 45.4%; Score 188.4; DB 6; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 2e-24;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 226 gtacatgcccgaataaagaaggaattttagatgttattcccatcttgaagaataat 285  
|||||  
Db 54 GTACATGGTCGATAAGAAAGCAATTGTAGATGTTAAATCCCATCTTGAAAGAAATAT 113  
QY 286 agtttaaatatttattgataaaataacaagtcaggtattatagtcgaagcaaaacataa 345  
|||||  
Db 114 AGTTTAAATATTATTGATAAAATAACCAAGTCAGGTATTATAGTCCAAAGCAAAACATAA 173  
QY 346 atttattgatgcgaatttcaattcagaataatttcaataactgattatcagctggtac 405  
|||||  
Db 174 ATTUATTCATCAAGTTTAAATTCAGAAATATTTCAATACTGATTATATCAGCTGGTAC 233  
QY 406 attgccgtag 415  
|||||  
Db 234 ATTGCCGTAG 243

Search completed: February 25, 2002, 12:51:36  
Job time: 2894 sec





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:03:22 ; Search time 142.95 Seconds  
(without alignments)  
2488.912 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagtttggttcattga.....cagctgtacatgcgtag 415

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 930621 seqs, 42862619 residues  
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_1101.\*  
1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	415	100.0	415	22	Right (5') border
2	380.8	91.8	416	22	Left (3') border f
3	191.4	46.1	5864	17	Plasmid pTCol113 T-
4	191.4	46.1	5865	22	Chimeric T-DNA of
5	191.4	46.1	7599	22	Nucleotide sequenc
6	188.8	45.5	5228	22	Plasmid pT5172gelt
7	188.8	45.5	6539	21	E. coli plasmid pt
8	188.8	45.5	6548	17	Plasmid pT5174 use
9	188.8	45.5	6548	18	Plasmid pT5172. C
10	188.8	45.5	6548	21	E. coli plasmid pt
11	188.8	45.5	7492	22	Plasmid pT5346. U

c 12	188.4	45.4	1303	17	AAT39337	Plasmid pT588 (Eco
c 13	188.4	45.4	3201	12	AAQ14529	pPS029 Bt ICP codi
c 14	188.4	45.4	4832	22	AAH25423	Nucleotide sequenc
c 15	188.4	45.4	4946	18	AAT59531	T-DNA of plasmid p
c 16	188.4	45.4	4946	22	AAH25422	Nucleotide sequenc
c 17	188.4	45.4	5349	19	AAV23239	T-DNA of pT524.
c 18	188.4	45.4	5864	17	AAT39339	Plasmid pTCol113 T-
c 19	188.4	45.4	5865	22	AAQ06990	Chimeric T-DNA of
c 20	188.4	45.4	7639	14	AAQ42160	Plasmid pPS0212 co
c 21	188.4	45.4	7639	14	AAQ42159	Plasmid pJD884 con
c 22	182.4	44.0	1037	11	AAQ04705	USP-Promoter-casse
c 23	182.4	44.0	1085	11	AAQ04703	Legumin-signalpept
c 24	182.4	44.0	1160	11	AAQ04706	USP-signalpeptide
c 25	179.4	43.2	1077	22	AAH25439	Right flanking reg
c 26	177.4	42.7	3201	12	AAQ15144	pVE36 Bt ICP codin
c 27	153	36.9	1186	13	AAQ25707	Chimeric neo gene
c 28	146	35.2	3153	21	AAZ29122	Plasmid DV131 comp
c 29	146	35.2	3336	21	AAZ29121	Plasmid DV130 comp
c 30	146	35.2	3694	21	AAZ29124	Plasmid DV133 used
c 31	146	35.2	3877	21	AAZ29123	Plasmid DV132 used
c 32	146	35.2	24593	6	AAH50226	Sequence of opine
c 33	146	35.2	24596	6	AAH50182	Complete nucleotid
c 34	107.6	25.9	936	22	AAF58252	Oligonucleotide D1
c 35	107.6	25.9	936	22	AAF58254	Oligonucleotide D1
c 36	107.6	25.9	936	22	AAF58257	Oligonucleotide D1
c 37	107.6	25.9	936	22	AAF58259	Oligonucleotide D2
c 38	107.6	25.9	936	22	AAF58262	Oligonucleotide D2
c 39	107.6	25.9	938	22	AAF58255	Oligonucleotide D1
c 40	106.4	25.6	936	22	AAF58252	Oligonucleotide D1
c 41	106.4	25.6	936	22	AAF58254	Oligonucleotide D1
c 42	106.4	25.6	936	22	AAF58257	Oligonucleotide D1
c 43	106.4	25.6	936	22	AAF58259	Oligonucleotide D2
c 44	106.4	25.6	936	22	AAF58262	Oligonucleotide D2
c 45	106.4	25.6	938	22	AAF58255	Oligonucleotide D1

ALIGNMENTS

RESULT 1  
AAD06997  
ID AAD06997 standard; DNA; 415 BP.  
AC AAD06997;  
XX  
DT 06-AUG-2001 (first entry)  
XX  
DE Right (5') border flanking region of elite event MS-B2.  
XX  
KW MS-B2 elite event; transgenic Brassica plant; transformation event;  
KW male-sterility gene; ds.  
XX  
OS Chimeric - Agrobacterium sp.  
OS Chimeric - Brassica sp.  
FH  
FH Key Location/Qualifiers  
FT misc\_feature 1..234  
FT /\*tag= a  
FT /note= "Corresponds to plant DNA"  
FT misc\_feature 235..415  
FT /\*tag= b  
FT /note= "Corresponds to T-DNA"  
XX  
XX WO200131042-A2.  
PN  
XX  
PD 03-MAY-2001.  
XX  
PF 26-OCT-2000; 2000WO-EP10680.  
XX  
PR 29-OCT-1999; 99US-0430497.  
XX  
PA (AVET ) AVENTIS CROPS SCIENCE NV.  
XX



```

RESULT 3
AAAT39339/c
ID AAAT39339 standard; DNA; 5864 BP.
XX
AC AAAT39339;
XX
DT 22-JAN-1997 (first entry)
XX
DE Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.
XX
KW Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;
KW transgenic plant; oilseed rape; canole; Brassica napus; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature complement (1..25)
FT /tag= a
FT /label= RB
FT /note= "right border of Agrobacterium T-DNA"
FT polyA_signal complement (98..330)
FT /tag= b
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT CDS complement (331..882)
FT /tag= c
FT /label= bar
FT /note= "region coding for phosphinothricin
FT acetyltransferase"
FT complement (883..2608)
FT /tag= d
FT /label= Pssu
FT /note= "promoter of Arabidopsis Rubisco small
FT subunit gene"
FT polyA_signal complement (2659..3031)
FT /tag= e
FT /label= 3'nos
FT /note= "region containing polyA signal of nopaline
FT synthase gene of Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /tag= f
FT /label= Barnase
FT /note= "Bacillus amyloliquefaciens barnase coding
FT region"
FT promoter complement (3368..4877)
FT /tag= g
FT /label= PTA29
FT /note= "promoter of stamen-specific PTA29 gene of
FT Nicotiana tabacum"
FT promoter 4924..5216
FT /tag= h
FT /label= Pnos
FT /note= "promoter of nopaline synthase gene of
FT Agrobacterium T-DNA"
FT CDS 5217..5489
FT /tag= i
FT /label= Barstar
FT /note= "region coding for barstar of Bacillus
FT amyloliquefaciens"
FT polyA_signal 5490..5765
FT /tag= j
FT /label= 3'g7
FT /note= "region containing polyA signal of gene 7
FT of Agrobacterium T-DNA"
FT misc_feature complement (5840..5864)
FT /tag= k
FT /label= LB
FT /note= "left border of Agrobacterium T-DNA"
XX
PN W09626283-A1.
XX
PD 29-AUG-1996.

```

```

XX 21-FEB-1996; 9GWO-EP00722.
PF
XX 21-FEB-1995; 95EP-0400364.
PR
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
PA
XX
PI Botterman J, Cornelissen M, Michiels F;
XX
DR WPI; 1996-402373/40.
XX
XX Prodn. of male sterile plants by transforming with a chimaeric
PT construct - comprising a male sterility DNA e.g. barnase and a
PT co-regulating gene, e.g. barstar, into the nuclear genome, useful
PT for generating hybrid cultivars
XX
PS Example 3; Page 33-3743-47; 56pp; English.
XX
CC Plasmid pTCO113 (AY39339) is a T-DNA vector containing a bar gene
CC under control of the PSSU promoter, a barnase gene under control
CC of the stamen-specific PTA29 promoter, and a barstar (co-regulatory)
CC gene under control of the Pnos promoter. 87% Of oilseed rape
CC plants regenerated after Agrobacterium-mediated transformation
CC using pTCO113 were male sterile. Barnase expression disturbed the
CC function of stamen cells leading to male sterility. Constitutive
CC expression of barstar counteracted any low level expression of
CC barnase in non-stamen tissue.
XX
SQ Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match 46.1%; Score 191.4; DB 17; Length 5864;
Best Local Similarity 99.5%; Pred. No. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 223 gatgtacatggccgataaagaaagcaattttagatgttaattcccatcttgaagaaa 282
|||||
Db 5812 GATGTACATGGTCGATAGAAAAGGCAATTGTAGATGTTAATTCCTTCATCTTGAAGAAA 5753
|||||

QY 283 tatagttaaattatttgataaaataacaagtcaggtattatagtcacgaagcaaaaa 342
|||||
Db 5752 TATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATATAGTCCACGCAAAAACA 5693
|||||

QY 343 taaattattgatgcaagtttaaattcagaataatttcaactgattatcatcagctgg 402
|||||
Db 5692 TAAATTTATTGATGCAAGTTTAAATTCAGAATAATTTCAATACTGATTATATATCAGCTGG 5633
|||||

QY 403 tacattgcccgtag 415
|||||
Db 5632 TACATTGCCGTAG 5620

RESULT 4
AA006990/c
ID AA006990 standard; DNA; 5865 BP.
XX
AC AA006990;
XX
DT 06-AUG-2001 (first entry)
XX
DE Chimeric T-DNA of plasmid pTCO113.
XX
KW T-DNA; plasmid pTCO113; transgenic Brassica plant; transformation event;
KW male-sterility gene; chimeric; tobacco; ds.
XX
OS Chimeric - Streptomyces hygroscopicus.
OS Chimeric - Arabidopsis thaliana.
OS Chimeric - Bacillus amyloliquefaciens.
OS Chimeric - Nicotiana tabacum.
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers

```



```
Qy 403 tacattgccgtag 415
Db 5633 TACATTGCCGTAG 5621

RESULT 5
AAF25320/c
ID AAF25320 standard; DNA; 7599 BP.
XX
AC AAF25320;
XX
DT 30-APR-2001 (first entry)
XX
DE Nucleotide sequence of a plasmid PGKB5.
XX
KW Plant promoter; root cell; root-specific expression; parasite resistance;
KW nematode resistance; fungal resistance; water stress; salt stress;
KW sugar content; nitrogen transport; ss.
XX
OS Synthetic.
XX
PN WO200100833-A1.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-FR01768.
XX
PR 25-JUN-1999; 99FR-0008185.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Hoffmann B, Mollier P, Pelletier G;
XX
DR WPI; 2001-102893/11.
XX
PT New constitutive plant promoter active specifically in roots, useful
PT for controlling expression of pest or drought resistance genes, and
PT related transgenic plants -
XX
PS Disclosure; Fig 9; 92pp; French.
XX
CC The present sequence represents a plasmid pCKB5. The plasmid contains
CC a plant promoter that directs expression of a selected sequence in
CC root calls at all stages of development of a plant. The plant promoter
CC is used to control expression of genes in a root-specific manner,
CC especially genes that provide resistance to parasites, pests (nematodes
CC or fungi), water and salt stress, or alter sugar content or nitrogen
CC transport. Fragments of the promoter are useful as probes or primers
CC to detect or amplify at least part of the promoter.
XX
SQ Sequence 7599 BP; 1972 A; 1938 C; 1937 G; 1752 T; 0 other;

Query Match 46.1%; Score 191.4; DB 22; Length 7599;
Best Local Similarity 99.5%; Pred. No. 1.4e-30;
Matches 192; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 223 gatgtacatggccgataaagaaagcaatttttagatgttaattcccatcttgaagaaa 282
Db 6771 GATGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTTGAAGAAA 6712
Qy 283 tatagtttaaatattattgataaaataacaagtcagggtattattagtcacagcaaaaaaca 342
Db 6711 TATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACA 6652
Qy 343 taaattattgatgcaagtttaaatccagaaatatttcaataactgattatcagctgg 402
Db 6651 TAAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATCAGCTGG 6592
Qy 403 tacattgccgtag 415
Db 6591 TACATTGCCGTAG 6579
```

```
RESULT 6
AAF86439/c
ID AAF86439 standard; DNA; 5228 BP.
XX
AC AAF86439;
XX
DT 25-JUN-2001 (first entry)
XX
DE Plasmid pTS172delta.
XX
KW Male sterile plant; RNAase inhibitor; plasmid pTS172delta; ds.
XX
OS Unidentified.
XX
PN WO200124616-A1.
XX
PD 12-APR-2001.
XX
PF 12-SEP-2000; 2000WO-JP06222.
XX
PR 30-SEP-1999; 99JP-0279307.
XX
PA (NISB ) JAPAN TOBACCO INC.
XX
PI Hamada K, Nakakido F;
XX
DR WPI; 2001-266212/27.
XX
PT Method for producing male sterile rice and maize by inserting RNAse
PT gene and RNAse inhibitor genes with promoters into the plant genome -
XX
PS Disclosure; Page 14-17; 29pp; Japanese.
XX
CC The present invention relates to a method for producing male sterile
CC plants. The method comprises inserting a promoter fragment upstream of an
CC RNAse gene and a second promoter, upstream of an RNAse inhibitor protein
CC gene and inserting it into the plant genome. The method is useful for
CC producing male sterile tobacco, lettuce and rapeseed plants, but
CC preferably rice and maize. The present sequence is a vector used in
CC the method of the present invention.
XX
SQ Sequence 5228 BP; 1384 A; 1307 C; 1263 G; 1274 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 5228;
Best Local Similarity 99.0%; Pred. No. 4.6e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 224 atgtacatggccgataaagaaagcaatttttagatgttaattcccatcttgaagaaaat 283
Db 5222 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATTCCTTGAAGCAAT 5163
Qy 284 atagtttaaatattattgataaaataacaagtcagggtattattagtcacagcaaaaaaat 343
Db 5162 ATAGTTTAAATATTATTGATATAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACAT 5103
Qy 344 aaatttattgatgcaagtttaaatccagaaatatttcaataactgattatcagctgg 403
Db 5102 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATATCAGCTGGT 5043
Qy 404 acattgccgtag 415
Db 5042 ACATTGCCGTAG 5031

RESULT 7
AAZ91097/c
ID AAZ91097 standard; DNA; 6539 BP.
XX
AC AAZ91097;
XX
DT 06-JUN-2000 (first entry)
```

XX DE E. coli plasmid pTS431 containing mutant barnase gene.  
 XX KW Male sterile plant; mutant barnase gene; anther-specific expression;  
 KW KW low fidelity PCR; primer; plant breeding; ss.  
 XX OS Synthetic.  
 XX PN WO200008176-A1.  
 XX PD 17-FEB-2000.  
 XX PF 03-AUG-1999; 99WO-JP04167.  
 XX PR 04-AUG-1998; 98JP-0220060.  
 XX PA (NISR) JAPAN TOBACCO INC.  
 XX PI Hamada K, Nakakido F;  
 XX WPI; 2000-195581/17.  
 XX PT Mutate barnase gene for efficient construction of plant transformants,  
 PT particularly male sterile plants free from any undesirable characters  
 PT by specifically expressing the gene alone in anther.  
 XX PS Example 3; Page 23-27; 30pp; Japanese.  
 XX CC The invention relates to the generation of male sterile plants by  
 CC the introduction of a mutant barnase gene (AA291095) for expression  
 CC specifically in the anther of a plant. This sequence represents the  
 CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains  
 CC the mutated barnase gene (AA291095) under control of the cauliflower  
 CC mosaic virus 35S promoter. The vector also contains a region of the  
 CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the  
 CC barnase gene to plants via an Agrobacterium tumefaciens host cell.  
 CC The transformed plant is used in plant breeding.  
 XX SQ Sequence 6539 BP; 1755 A; 1578 C; 1519 G; 1687 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6539;  
 Best Local Similarity 99.0%; Pred. No. 4.7e-30;  
 Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 224 atgtacatgccgatagaagaagcaattttagatgttaattcccatcttgaagaagaat 283  
 Db 6533 ACGTACATGGTCGATAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAGAAT 6474

Qy 284 atagtttaaatatttattgataaaatacaagtcaggtattatagtcacgaacaaacat 343  
 Db 6473 ATAGTTTAAATATTATTGATTAATAAACAAGTCAGGTATTATAGTCCAGCAAAACAT 6414

Qy 344 aaattttatgacgaagttaattcaaatcagaataatttcaataactgattatcagctggt 403  
 Db 6413 AAATTATTATGACGAAGTTTAAATTCAGAAATATTTCATAAATGATTATATCAGCTGGT 6354

Qy 404 acattgcccgtag 415  
 Db 6353 ACATTGCGGTAG 6342

RESULT 8  
 AAT39336/c  
 ID AAT39336 standard; DNA; 6548 BP.  
 XX AC AAT39336;  
 XX AC AAT39336;  
 DT 22-JAN-1997 (first entry)  
 XX DE Plasmid pTS174 used to obtain male sterile rice.  
 XX KW Plasmid pTS174; male sterile; barnase; ribonuclease; transgenic plant;

KW rice; Oryza sativa; ds; cyclic.  
 XX Synthetic.  
 XX PH Key Location/Qualifiers  
 FT misc\_feature 1..2003  
 FT /\*tag= a  
 FT /label= Vector  
 FT /note= "pUC19 derived vector sequences"  
 FT complement (2019..2283)  
 FT /\*tag= b  
 FT /label= 3'nos  
 FT /note= "region containing polyadenylation signal  
 FT nopaline synthase gene of Agrobacterium  
 FT T-DNA"  
 FT CDS complement (2284..2624)  
 FT /\*tag= c  
 FT /label= Barnase  
 FT /product= Bacillus amyloliquefaciens barnase  
 FT complement (2625..4313)  
 FT /\*tag= d  
 FT /label= PEI  
 FT /function= promoter of the stamen-specific EI gene  
 FT of rice  
 FT promoter 4336..5710  
 FT /\*tag= e  
 FT /label= P35S  
 FT /function= 35S promoter of cauliflower mosaic virus  
 FT 5711..6262  
 FT /\*tag= f  
 FT /label= bar  
 FT /product= phosphinothricin acetyltransferase  
 FT 6263..6456  
 FT /\*tag= g  
 FT /label= 3'g7  
 FT /function= region containing polyadenylation signal  
 XX WO9626283-A1.  
 XX 29-AUG-1996.  
 XX 21-FEB-1996; 96WO-BP00722.  
 XX 21-FEB-1995; 95EP-0400364.  
 XX (PLBZ) PLANT GENETIC SYSTEMS NV.  
 XX Botterman J, Cornelissen M, Michiels F;  
 XX WPI; 1996-402373/40.  
 XX Prodn. of male sterile plants by transforming with a chimaeric  
 XX construct - comprising a male sterility DNA e.g. barnase and a  
 XX co-regulating gene, e.g. Barstar, into the nuclear genome, useful  
 XX for generating hybrid cultivars  
 PS Example 1; Page 33-37; 56pp; English.  
 XX Plasmid pTS174 (AAT39336) contains Bacillus barnase DNA under control  
 CC of the stamen-specific PEI promoter. Embryogenic callus from rice  
 CC cv. Koshihikari was transformed with pTS174 alone or with pTS88  
 CC (see also AAT39337), a plasmid contg. Barstar DNA under control of a  
 CC 35S promoter. With pTS174 alone, 1 male sterile line was recovered  
 CC from 48 electroporation cuvettes. With both plasmids, 7 normal  
 CC male sterile lines were recovered from 40 cuvettes. Barnase  
 CC expression disturbed the function of stamen cells leading to male  
 CC sterility. Constitutive expression of Barstar counteracted any low  
 CC level expression of barnase in non-stamen tissue.  
 XX SQ Sequence 6548 BP; 1757 A; 1578 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 17; Length 6548;



Best Local Similarity 99.0%; Pred. No. 4.7e-30;		Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Qy	224 atgtacatggccgataaagaagcaattttagatgttaattcccatcttgaagaataa 283		
Db	6542 ACGTACATGGTCGATAAGAAAGCAATTTAGATGTTAATTCCTTCATCTTGAAGAAAT 6483		
Qy	284 atagttaaataattattgataataaatacaagtcagggtattatagtcacaagcaaaaaaat 343		
Db	6482 ATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACAT 6423		
Qy	344 aaattattgatgaagtttaattcagaataatttcataactatattatcagctggt 403		
Db	6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6363		
Qy	404 acattgcccgtag 415		
Db	6362 ACATTGCCGTAG 6351		
RESULT 9			
AAAT61394/c			
ID	AAAT61394 standard; DNA; 6548 BP.		
XX	AAAT61394;		
AC	AAAT61394;		
XX	07-MAY-1997 (first entry)		
DT			
XX	Plasmid pTS172.		
XX	Transgenic plant; poly-(ADP-ribose) polymerase inhibitor; PARP;		
KW	niacinamide; Agrobacterium; T-DNA; male sterile; barnase;		
KW	ribonuclease; RNase; cereal; wheat; Triticum aestivum;		
KW	plasmid pTS172; ds.		
XX			
OS	Chimeric Agrobacterium sp.;		
OS	Chimeric Oryza sativa;		
OS	Chimeric cauliflower mosaic virus.		
XX			
FH	Key Location/Qualifiers		
FT	3'UTR complement (2019..2288)		
FT	/tag= a		
FT	/label= 3'nos		
FT	/note= "3' untranslated region contg. the poly-A		
FT	signal of Agrobacterium T-DNA nopaline		
FT	synthase gene"		
FT	complement (2289..2624)		
FT	/tag= b		
FT	/product= barnase		
FT	complement (2625..4313)		
FT	/tag= c		
FT	/label= PE1		
FT	/note= "promoter region of rice E1 gene"		
FT	complement (4336..5710)		
FT	/tag= d		
FT	/label= P35S		
FT	/note= "35S promoter region of cauliflower mosaic		
FT	virus"		
FT	5711..6262		
FT	/tag= e		
FT	/label= Bar		
FT	/note= "phosphinothricin acetyltransferase"		
FT	6243..6496		
FT	/tag= f		
FT	/label= 3'g7		
FT	/note= "3' untranslated region contg. the poly-A		
FT	signal of gene 7 of Agrobacterium T-DNA"		
XX			
PN	EP757102-A1.		
XX			
PD	05-FEB-1997.		
XX			
PF	04-AUG-1995; 95EP-0401844.		
XX	04-AUG-1995; 95EP-0401844.		
XX	(PLBZ ) PLANT GENETIC SYSTEMS NV.		
PI	De Block M;		
XX	WPI; 1997-111050/11.		
DR			
XX	Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase		
PT	inhibitor - reduces the cultured cells response to stress and		
PT	reduces metabolism		
XX			
PS	Example 2; Page 17-20; 25pp; English.		
XX	Plasmid pTS172 (AAT61394) contains the barnase coding sequence under		
CC	control of the rice E1 gene stamen-specific promoter and a		
CC	phosphinothricin acetyltransferase coding sequence under control of		
CC	the CamV 35S promoter. Plasmid pTS172 and plasmid pTS772 (see also		
CC	AAAT61395) were used to transform wheat Spring variety Pavon calli via		
CC	particle bombardment. Some calli were treated with the poly-(ADP-		
CC	ribose) polymerase inhibitor niacinamide before, or before and		
CC	after, bombardment. Healthy, male sterile plants were regenerated		
CC	only from bombarded calli that were treated with niacinamide. This		
CC	was believed to be due to more faithful expression characteristics		
CC	of the integrated stamen-selective batnase gene in these calli		
CC	and regenerated shoots. For plants transformed with pTS172,		
CC	foreign DNA was stably incorporated in the wheat genome in 2-3		
CC	copies.		
XX			
SQ	Sequence 6548 BP; 1756 A; 1579 C; 1523 G; 1690 T; 0 other;		
Query Match 45.5%; Score 188.8; DB 18; Length 6548;		Best Local Similarity 99.0%; Pred. No. 4.7e-30;	
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	224 atgtacatggccgataaagaagcaattttagatgttaattcccatcttgaagaataa 283		
Db	6542 ACGTACATGGTCGATAAGAAAGCAATTTAGATGTTAATTCCTTCATCTTGAAGAAAT 6483		
Qy	284 atagttaaataattattgataataaatacaagtcagggtattatagtcacaagcaaaaaaat 343		
Db	6482 ATAGTTTAAATATTATTGATAAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACAT 6423		
Qy	344 aaattattgatgaagtttaattcagaataatttcataactatattatcagctggt 403		
Db	6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAACTGATTATATCAGCTGGT 6363		
Qy	404 acattgcccgtag 415		
Db	6362 ACATTGCCGTAG 6351		
RESULT 10			
AAZ91096/c			
ID	AAZ91096 standard; DNA; 6548 BP.		
XX	AAZ91096;		
XX	06-JUN-2000 (first entry)		
DT			
DE	E. coli plasmid pTS172 containing synthetic barnase gene.		
XX			
KW	Male sterile plant; mutant barnase gene; anther-specific expression;		
KW	low fidelity PCR; primer; plant breeding; ss.		
XX			
OS	Synthetic.		
XX			
PN	WO200008176-A1.		
XX			
PD	17-FEB-2000.		
XX			

```

PF 03-AUG-1999; 99WO-JP04167.
XX
PR 04-AUG-1998; 98JP-0220060.
XX
PA (NISB ) JAPAN TOBACCO INC.
XX
PI Hamada K, Nakakido F;
XX WPI; 2000-195581/17.
XX
XX Mutate barnase gene for efficient construction of plant transformants,
PT particularly male sterile plants free from any undesirable characters
PT by specifically expressing the gene alone in anther .
XX
PS Example 3; Page 19-23; 30pp; Japanese.
XX
CC The invention relates to the generation of male sterile plants by
CC the introduction of a mutant barnase gene (AAZ91095) for expression
CC specifically in the anther of a plant. This sequence represents the
CC E. coli/Agrobacterium shuttle vector plasmid pTS172 which contains
CC the synthetic barnase gene (AAZ91094) under control of the cauliflower
CC mosaic virus 35S promoter. The vector also contains a region of the
CC Agrobacterium T-DNA gene 7. The vector is used for transmitting the
CC barnase gene to plants via an Agrobacterium tumefaciens host cell.
XX The transformed plant is used in plant breeding.
XX
SQ Sequence 6548 BP; 1756 A; 1579 C; 1523 G; 1690 T; 0 other;

Query Match 45.5%; Score 188.8; DB 21; Length 6548;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggcgcgataagaagaagccaatttgcagatgtaattcccatcttgaagaagaat 283
DB 6542 ACGTACATGGTGGATAGAAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAGAAAT 6483
QY 284 atagtttaaatatttattgataaaatacaagtcaggtattattatgcagtcggt 403
DB 6482 ATAGTTTAAATATTATTGTGATAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACAT 6423
QY 344 aaattattatgcaggttttaattcagaataatttcaataactgattatcagctggt 403
DB 6422 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAAGTATTATATCAGCTGGT 6363
QY 404 acattgcgctag 415
DB 6362 ACATTGCCGTAG 6351

RESULT 11
AAF86441/C
ID AAF86441 standard; DNA; 7492 BP.
XX
XX AAF86441;
AC
XX
DT 25-JUN-2001 (first entry)
XX
XX Plasmid pTS346.
DE
XX Male sterile plant; RNAase inhibitor; plasmid pTS346; ds.
XX
XX Unidentified.
OS
XX WO200124616-A1.
PN
XX 12-APR-2001.
PD
XX 12-SEP-2000; 2000WO-JP06222.
PF
XX 30-SEP-1999; 99JP-0279307.
PR
XX (NISB ) JAPAN TOBACCO INC.
PA

```

```

XX
PI Hamada K, Nakakido F;
XX
DR WPI; 2001-266212/27.
XX
PT Method for producing male sterile rice and maize by inserting RNase
PT gene and RNase inhibitor genes with promoters into the plant genome -
XX Disclosure; Page 19-23; 29pp; Japanese.
XX
XX The present invention relates to a method for producing male sterile
XX plants. The method comprises inserting a promoter fragment upstream of an
XX RNase gene and a second promoter, upstream of an RNase inhibitor protein
XX gene and inserting it into the plant genome. The method is useful for
XX producing male sterile tobacco, lettuce and rapeseed plants, but
XX preferably rice and maize. The present sequence is a vector used in
XX the method of the present invention..
XX
SQ Sequence 7492 BP; 1987 A; 1801 C; 1752 G; 1952 T; 0 other;

Query Match 45.5%; Score 188.8; DB 22; Length 7492;
Best Local Similarity 99.0%; Pred. No. 4.7e-30;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 224 atgtacatggcgcgataagaagaagccaatttgcagatgtaattcccatcttgaagaagaat 283
DB 7486 ACGTACATGGTGGATAGAAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAAGAAAT 7427
QY 284 atagtttaaatatttattgataaaatacaagtcaggtattattatgcagtcggt 403
DB 7426 ATAGTTTAAATATTATTGTGATAAATAACAAGTCAGGTATTATAGTCCCAAGCAAAAACAT 7367
QY 344 aaattattatgcaggttttaattcagaataatttcaataactgattatcagctggt 403
DB 7366 AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCATAAAGTATTATATCAGCTGGT 7307
QY 404 acattgcgctag 415
DB 7306 ACATTGCCGTAG 7295

RESULT 12
AAT39337/C
ID AAT39337 standard; DNA; 1303 BP.
XX
XX AAT39337;
AC
XX
DT 22-JAN-1997 (first entry)
XX
XX Plasmid pTS88 (EcoRI-HindIII fragment).
DE
XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;
KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH misc_feature 1..35
FT /*tag= a
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
FT 36..694
FT /*tag= b
FT /label= p35S
FT /function= 35S promoter of cauliflower mosaic virus
FT strain CM1841
FT CDS 695..967
FT /*tag= c
FT /label= barstar
FT /product= Bacillus amyloliquefaciens barstar
FT 968..1287
FT polyA_signal
FT /*tag= d

```

```

FT /label= 3'g7
FT /function= region containing polyadenylation signal
FT of gene 7 og Agrobacterium T-DNA
FT misc_feature 1288..1303
FT /*tag= e
FT /label= pGEM2
FT /note= "polylinker of pGEM2"
XX
XX WO9626283-A1.
XX
XX 29-AUG-1996.
XX
XX 21-FEB-1996; 96WO-EP00722.
XX
XX 21-FEB-1995; 95EP-0400364.
XX
XX (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX Botterman J, Cornelissen M, Michiels F;
XX
XX WPI; 1996-402373/40.
XX
XX Prodn. of male sterile plants by transforming with a chimaeric
XX construct - comprising a male sterility DNA e.g. barnase and a
XX co-regulating gene, e.g. barstar, into the nuclear genome, useful
XX for generating hybrid cultivars
XX
XX Example 1; Page 38; 56pp; English.
XX
XX The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains
XX barstar DNA under control of a 35S promoter. The plasmid was
XX used with pTS174 (see also AAT39336) contg. barnase DNA under
XX control of the stamen-specific promoter El to produce male sterile
XX rice cv. Kochihibiki transgenic plants, and with plasmid pVEL36
XX (see also AAT39338) contg. barnase DNA under control of the stamen-
XX specific PCAS5 promoter to produce male sterile maize plants.
XX Expression of barnase (a ribonuclease) in the stamen leads to male
XX sterility. Constitutive expression of barstar counteracts possible
XX low level expression of barnase DNA in non-stamen tissue.
XX
XX Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;
XX
XX
XX Query Match 45.4%; Score 188.4; DB 17; Length 1303;
XX Best Local Similarity 99.5%; Pred. No. 5e-30;
XX Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 226 gtacatgcccgaataagaagcaattttagatgttaattcccatcttgaagaataat 285
XX 1287 GTACATGTCGTGATAAGAAAGGCAATTTGTAGATGTTAATTCCTTCTTGAAGAATAAT 1228
XX
XX 286 agtttaaatattattgataaaataacaagtcaggtattattagtcacagcaaaaacataa 345
XX 1227 AGTTTAAATATTATTGATAAAATAACAAAGTCAGGTATTATTAGTCCAGCAAAAACATAA 1168
XX
XX 346 atttattgatgcaagtctaaatttcagaataatttccaataactgatttatatcagctgggtac 405
XX 1167 ATTTATTGATGCAAGCTTTAAATTCAGAAATATTTCATAAAGTATTATATATCAGCTGGTAC 1108
XX
XX 406 attgcccgtag 415
XX 1107 ATTGCCGTAG 1098
XX
XX
XX RESULT 13
XX AAQ14529/c
XX ID AAQ14529 standard; DNA; 3201 BP.
XX
XX AC
XX AAQ14529;
XX
XX 27-JAN-1992 (first entry)
XX
XX pPS029 Bt ICP coding sequence.

```

```

XX
XX Bacillus thuringiensis; insecticidal crystal protein; ICP;
XX deletion; ss.
XX
XX Synthetic.
XX
XX WO9116432-A.
XX
XX 31-OCT-1991.
XX
XX 17-APR-1991; 91WO-EP00733.
XX
XX 18-APR-1990; 90EP-0401055.
XX
XX (PLAN-) PLANT GENETIC SYST.
XX
XX Cornelissen M, Soetaert P, Stam M, Dockx J;
XX
XX WPI; 1991-339820/46.
XX
XX Modified Bacillus thuringiensis insecticidal crystal protein
XX genes - having A and T sequences changed to G and C sequences
XX encoding same amino acids, for increased expression levels
XX
XX Disclosure; Fig 6(c); 78pp; English.
XX
XX "n" in the sequence refers to not known nucleotides.
XX pPS029 is identical to pVE36 (AAQ15144), but carries both the amino-
XX terminal modification and the internal modification of the Bt ICP
XX coding sequence.
XX See also AAQ14529, AAQ15142-44.
XX
XX Sequence 3201 BP; 880 A; 710 C; 720 G; 886 T; 5 other;
XX
XX
XX Query Match 45.4%; Score 188.4; DB 12; Length 3201;
XX Best Local Similarity 99.5%; Pred. No. 5.3e-30;
XX Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 226 gtacatgcccgaataagaagcaattttagatgttaattcccatcttgaagaataat 285
XX 3141 GTACATGTCGTGATAAGAAAGGCAATTTGTAGATGTTAATTCCTTCTTGAAGAATAAT 3082
XX
XX 286 agtttaaatattattgataaaataacaagtcaggtattattagtcacagcaaaaacataa 345
XX 3081 AGTTTAAATATTATTGATAAAATAACAAAGTCAGGTATTATTAGTCCAGCAAAAACATAA 3022
XX
XX 346 atttattgatgcaagtctaaatttcagaataatttccaataactgatttatatcagctgggtac 405
XX 3021 ATTTATTGATGCAAGCTTTAAATTCAGAAATATTTCATAAAGTATTATATATCAGCTGGTAC 2962
XX
XX 406 attgcccgtag 415
XX 2961 ATTGCCGTAG 2952
XX
XX
XX RESULT 14
XX AAH25423
XX ID AAH25423 standard; DNA; 4832 BP.
XX
XX AC
XX AAH25423;
XX
XX 22-AUG-2001 (first entry)
XX
XX Nucleotide sequence of plasmid pTHW118.
XX
XX Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX fertility restorer gene; barstar gene; ss.
XX
XX Synthetic.
XX
XX Streptomyces hygroscopicus.
XX
XX Arabidopsis thaliana.
XX
XX Bacillus amyloliquefaciens.

```



```
FT CDS complement (331..882)
FT /*tag= c
FT /label= Bar
FT /product= phosphinothricin acetyltransferase
FT promoter complement (883..2608)
FT /*tag= d
FT /label= PSSU
FT /note= "promoter region of Rubisco small subunit
FT gene of Arabidopsis thaliana"
FT 3'UTR complement (2658..3031)
FT /*tag= e
FT /label= 3'nos
FT /note= "3'untranslated region contg. the poly-A
FT signal of the nopaline-synthase gene of
FT Agrobacterium T-DNA"
FT CDS complement (3032..3367)
FT /*tag= f
FT /label= Barnase
FT /product= barnase
FT complement (3368..4876)
FT /*tag= g
FT /label= PTA29
FT /note= "promoter region of tobacco PTA29 gene"
FT misc_RNA complement (4822..4946)
FT /*tag= h
FT /label= LB
FT /note= "T-DNA left border"
FT
FT
XX EP757102-A1.
XX
XX
XX 05-FEB-1997.
XX
XX 04-AUG-1995; 95EP-0401844.
XX
XX 04-AUG-1995; 95EP-0401844.
XX
XX (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX
XX De Block M;
XX
XX WPI; 1997-111050/11.
XX
XX Prodn. of transgenic plants using a poly-(ADP-ribose) polymerase
XX inhibitor - reduces the cultured cells response to stress and
XX reduces metabolism
XX
XX Example 3; Page 13-16; 25pp; English.
XX
XX Plasmid pTHW107 is a vector carrying T-DNA (AAV59531) comprising a
XX barnase coding sequence under control of the tobacco TA29 gene
XX stamen-specific promoter and a phosphinothricin acetyltransferase
XX coding sequence under control of an Arabidopsis Rubisco small
XX subunit gene promoter. Oilseed rape hypocotyl explants were
XX infected with Agrobacterium tumefaciens C58ClRif carrying vector
XX pTHW107 and helper Ti plasmid pMP60. In some treatments, the
XX hypocotyls were treated with the poly-(ADP-ribose) polymerase
XX inhibitor niacinamide (250 mg/l) 4 days prior to infection.
XX Plants regenerated from niacinamide-treated transformed calli
XX had a low copy number and displayed less variation in the
XX expression profile of the transgenes.
XX
XX Sequence 4946 BP; 1569 A; 891 C; 963 G; 1523 T; 0 other;

Query Match 45.4%; Score 188.4; DB 18; Length 4946;
Best Local Similarity 99.5%; Pred. NO. 5.5e-30;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatgccgataaagaaggcaattgttagatttaattcccatcttgaagaataat 285
Db 54 gtacatggtcataagaagaaggcaattgttagatttaattcccatcttgaagaataat 113
Qy 286 agtttaaatattattgataaaataacaagtcaggtattattagtcacaagcaaaacataa 345
```

```
Db 114 agtttaaatattattgataaaataacaagtcaggtattattagtcacaagcaaaacataa 173
Qy 346 atttattgatgcaagttttaaaattcagaataatttcaataaactgattatcagctggtac 405
Db 174 atttattgatgcaagttttaaaattcagaataatttcaataaactgattatcagctggtac 233
Qy 406 attgccgtag 415
Db 234 attgccgtag 243

Search completed: February 25, 2002, 12:55:46
Job time: 3144 sec
```

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:03:22 ; Search time 80.64 seconds  
(without alignments)  
1165.530 Million cell updates/sec

Title: US-09-698-903B-8

Perfect score: 415

Sequence: 1 gtcgagtttggtgttcataga.....cagctggtacattgcgtag 415

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/1na/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/1na/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/1na/PTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/1na/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	191.4	46.1	5864	3	US-08-894-440-4
C 2	188.8	45.5	6548	3	US-08-894-440-1
C 3	188.8	45.5	6548	3	US-08-817-188-2
C 4	188.8	45.5	7811	2	US-08-549-680A-5
C 5	188.4	45.4	1303	3	US-08-894-440-2
C 6	188.4	45.4	3200	1	US-08-453-104-23
C 7	188.4	45.4	3200	2	US-08-694-824-23
C 8	188.4	45.4	4946	3	US-08-817-188-1
C 9	188.4	45.4	5560	3	US-08-817-188-5
C 10	188.4	45.4	7566	2	US-08-232-016-23
C 11	188.4	45.4	7639	2	US-08-232-016-22
C 12	186.8	45.0	5864	3	US-08-894-440-4
C 13	177	42.7	3201	1	US-08-453-104-22
C 14	177	42.7	3201	2	US-08-694-824-22
C 15	153	36.9	1186	1	US-08-064-121-2
C 16	153	36.9	1186	1	US-08-478-015-2
C 17	153	36.9	1186	3	US-08-475-975-2
C 18	153	36.9	1186	3	US-09-084-889-2
C 19	146	35.2	3153	4	US-09-080-625-3
C 20	146	35.2	3336	4	US-09-080-625-2
C 21	146	35.2	3694	4	US-09-080-625-5
C 22	146	35.2	3877	4	US-09-080-625-4
C 23	143	34.5	24595	6	5428147-1
C 24	44.8	10.8	8654	1	US-08-920-812-6
C 25	44.8	10.8	8654	1	US-08-920-827-6
C 26	44.8	10.8	8654	1	US-08-921-177-6
C 27	44.8	10.8	8654	1	US-08-362-577C-6

28	44.8	10.8	8654	2	US-08-920-828-6	Sequence 6, Appl
C 29	44.6	10.7	5526	3	US-08-751-359-21	Sequence 21, Appl
C 30	43.6	10.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 31	43	10.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 32	42.6	10.3	8920	2	US-08-446-855A-1	Sequence 1, Appl
C 33	42.6	10.3	8920	4	US-09-150-741-1	Sequence 1, Appl
C 34	42	10.1	1316	2	US-08-871-924A-1	Sequence 1, Appl
C 35	40.4	9.7	1588	3	US-09-058-489-45	Sequence 45, Appl
C 36	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appl
C 37	39.8	9.6	15397	2	US-08-673-768-1	Sequence 1, Appl
C 38	39.8	9.6	24595	6	5428147-1	Patent No. 5428147
C 39	39.4	9.5	80595	4	US-09-078-294-3	Sequence 3, Appl
C 40	39.2	9.4	658	4	US-08-998-416-595	Sequence 595, App
C 41	38.8	9.3	636	4	US-08-998-416-1137	Sequence 1137, Ap
C 42	38.8	9.3	660	1	US-07-991-867B-32	Sequence 32, Appl
C 43	38.8	9.3	660	1	US-08-107-755A-32	Sequence 32, Appl
C 44	38.8	9.3	660	2	US-08-544-332-32	Sequence 32, Appl
C 45	38.8	9.3	1511	1	US-07-991-867B-8	Sequence 8, Appl

#### ALIGNMENTS

RESULT 1  
US-08-894-440-4/c  
; Sequence 4, Application US/08894440  
; Patent No. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894,440  
; CURRENT FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 5864  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
; OTHER INFORMATION: plasmid pTCO113  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((1)..(25))  
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((98)..(330))  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((331)..(882))  
; OTHER INFORMATION: region coding for phosphinothricin acetyl  
; OTHER INFORMATION: transferase (bar)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((883)..(2608))  
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of  
; OTHER INFORMATION: Arabidopsis (Pssu)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((2659)..(3031))  
; OTHER INFORMATION: region containing polyadenylation signal of  
; OTHER INFORMATION: napaline synthase gene of Agrobacterium T-DNA  
; OTHER INFORMATION: (3'nos)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: Complement((3032)..(3367))  
; OTHER INFORMATION: region coding for barnase of Bacillus  
; OTHER INFORMATION: amyloliquefaciens  
; FEATURE:





```

; LOCATION: Complement((2019)..(2288))
; OTHER INFORMATION: 3' nos: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2289)..(2624))
; OTHER INFORMATION: barnase: region coding for barnase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2625)..(4313))
; OTHER INFORMATION: PEI: promoter region of EI gene of rice
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4336)..(5170)
; OTHER INFORMATION: P35S: 35S promoter region of Cauliflower Mosaic
; OTHER INFORMATION: Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5711)..(6262))
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6263)..(6496)
; OTHER INFORMATION: 3'g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; US-08-817-188-2

```

Query Match	45.5%	Score 188.8;	DB 3;	Length 6548;
Best Local Similarity	99.0%	Pred. No. 1.1e-36;		
Matches 190;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY 224	atgtacatggccgataagaagaagcaatttgtagatgttaattcccatcttgaagaagt	283		
Db				
6542	ACGTACATGTCGATAGAAAAGGCAATTTGTAGATGTTAATTCCCATCTTGAAGGAAT	6483		
QY 284	atagtttaaatattttattgataaaataacaagtccaggtattatagttccaaagcaaaaacat	343		
Db				
6482	ATAGTTTAAATATTATTATGATAAAATACCAAGTCAGGTATTATAGTCCAAGCAAAAACAT	6423		
QY 344	aaatttattgatgcgaagtttcaaatcagaagaattttcaataaactgattatatacagctggt	403		
Db				
6422	AAATTTATTCATGCAAGTTTAAATTCAGAAATATTTCCAACTCAATATATATCATCAGCTGCT	6363		
QY 404	acattggcgtag	415		
Db				
6362	ACATTGCCGTAG	6351		

```

RESULT      4
US-85-549-680A-5
, Sequence 5, Application US/08549680A
, Patent No. 5962768
, GENERAL INFORMATION:
, APPLICANT: CORNELIENS, MARCUS
, APPLICANT: REYNAERTS, ARLETTE
, APPLICANT: GOSSELE, VERONIQUE
, APPLICANT: VAN AKRSSEN, ROEL
, TITLE OF INVENTION: MARKER GENE
, NUMBER OF SEQUENCES: 6
, CORRESPONDENCE ADDRESS:
, ADDRESSEE: Birch, Stewart, Kolb
, STREET: P.O. Box 747
, CITY: Falls Church
, STATE: Virginia
, COUNTRY: USA
, ZIP: 22040-0747
, COMPUTER READABLE FORM:
, MEDIUM TYPE: Floppy disk
, COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,680A
; FILING DATE: 16 JANUARY 1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: SVENSSON, LEONARD R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-0111P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7811 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (synthetic)
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: 1..7811
; OTHER INFORMATION: /label= vector ptrVA3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 194..218
; OTHER INFORMATION: /note= "T-DNA right border"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 484..684
; OTHER INFORMATION: /note= "the 3' end formation and
; OTHER INFORMATION: polyadenylation region of T-DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (729..1340)
; OTHER INFORMATION: /note= "the aac(6') coding
; OTHER INFORMATION: sequence"
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1341..1756
; OTHER INFORMATION: /label= 35S promoter
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3001..3023
; OTHER INFORMATION: /note= "T-DNA left border
; OTHER INFORMATION: sequences"
; PS-08-549-680A-5

```

		Query Match	45.5%; Score 188.8; DB 2; Length 7811;	
		Best Local Similarity	99.0%; Pred. No. 1.2e-36;	
		Matches 190; Conservative	0; Mismatches 2; Indels 0; Gaps	0;
QY	224	atgtacatgcccgcataagaaaaggcaattttagatgttaattcccatcttgaaagaaat	283	
Db	438	ACGTACATGGTCGATAAGAAAGCAATTCTGATGTTAATCCCATCTTGAAGAANAAT	497	
QY	284	atagttaaataattattttgataaaaaacgaagtcagggtatttatagtcaccaagcAAAAACAAT	343	
Db	498	ATAGTTTTAAAATATTATTGATAAAAATACAAGTCAGGTATTATAGTCCAAGCAAAAACAAT	557	
QY	344	aaattattatgacgaagttttaattcgaaatatattcoataactgatttatatcacgctgg	403	
Db	558	AAATTTTATGCATCGAAGTTTAATTCAGAAAATATTTCAAATAACTGATTATATCACGCTG	617	
QY	404	acattgccgtag	415	
Db	618	ACATTGCCGTAG	629	
				RESULT 5



GENERAL INFORMATION:  
APPLICANT: CORNELISSEN, Marc  
APPLICANT: SOETAERT, Piet  
APPLICANT: STAM, Maïke  
APPLICANT: DOCKX, Jan  
TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
TITLE OF INVENTION: IN PLANT CELLS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/694,824  
FILING DATE: 09-AUG-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,869  
FILING DATE: 16-DEC-1992  
APPLICATION NUMBER: GB 90401055.0  
FILING DATE: 18-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Rea, Teresa S  
REGISTRATION NUMBER: 30,427  
REFERENCE/DOCKET NUMBER: 010830-032  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3200 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2078..2082  
OTHER INFORMATION: /note= "Nucleotides 2078-2082  
OTHER INFORMATION: wherein N is not known."  
US-08-694-824-23

Query Match 45.4%; Score 188.4; DB 2; Length 3200;  
Best Local Similarity 99.5%; Pred. No. 1.3e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 226 gtacatggccgataagaagcaattgttagatgttaattcccatcttgaaagaatat 285  
|||||  
Db 3141 GTACATGGTCGATAGAAAAGCAATTTGTAGATGTTAATCCCATCTTGAAGAAATAT 3082  
|||||  
Qy 286 agtttaaatatttattgataaaataacaagtcaggtattatagtcgaagcaaaaacataa 345  
|||||  
Db 3081 AGTTTAAATATTATTATGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACATAA 3022  
|||||  
Qy 346 atttattgatcgaagttaaataatttcagaataatttcataactgattatatacagctgggtac 405  
|||||  
Db 3021 ATTTATTGATCAAGTTTAAATTTCAAGAAATATTTCAATAACTGATTATATCAGCTGGTAC 2962  
|||||  
Qy 406 attgccgtag 415  
|||||  
Db 2961 ATTGCCGTAG 2952

RESULT 8  
US-08-817-188-1  
Sequence 1, Application US/08817188  
Patent No. 6074876  
GENERAL INFORMATION:  
APPLICANT: DE BLOCK, MARC  
TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR  
FILE REFERENCE: 2121-0127P  
CURRENT APPLICATION NUMBER: US/08/817,188  
CURRENT FILING DATE: 1997-05-15  
EARLIER APPLICATION NUMBER: PCT/EP96/03366  
EARLIER FILING DATE: 1996-07-31  
EARLIER APPLICATION NUMBER: EP 95401844.6  
EARLIER FILING DATE: 1995-08-04  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 4946  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of  
OTHER INFORMATION: plasmid pTHW107  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((1)..(25))  
OTHER INFORMATION: T-DNA right border (RB)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((97)..(330))  
OTHER INFORMATION: 3'g7; 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((331)..(882))  
OTHER INFORMATION: bar; region coding for phosphinotricin acetyl  
OTHER INFORMATION: transferase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((883)..(2608))  
OTHER INFORMATION: promoter region of Rubisco small subunit gene of  
OTHER INFORMATION: Arabidopsis thaliana (PSSU)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((2658)..(3031))  
OTHER INFORMATION: 3' nos; 3' untranslated region containing the  
OTHER INFORMATION: polyadenylation signal of the nopaline synthase  
OTHER INFORMATION: gene of Agrobacterium T-DNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3032)..(3367))  
OTHER INFORMATION: barnase; region coding for barnase  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((3368)..(4876))  
OTHER INFORMATION: PTA29; promoter region of TA29 gene of Nicotiana  
OTHER INFORMATION: tabacum  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: Complement((4922)..(4946))  
OTHER INFORMATION: LB: T-DNA left border  
US-08-817-188-1

Query Match 45.4%; Score 188.4; DB 3; Length 4946;  
Best Local Similarity 99.5%; Pred. No. 1.4e-36;  
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 226 gtacatggccgataagaagcaattgttagatgttaattcccatcttgaaagaatat 285  
|||||  
Db 54 gtacatggctcgataagaagcaattgttagatgttaattcccatcttgaaagaatat 113

```

QY 286 agttaaatattattgataaaataacaagtcagggtattattatagtcacgaacaaacataa 345
Db 114 agttaaatattattgataaaataacaagtcagggtattattatagtcacgaacaaacataa 173
QY 346 attattatgcgaagttaaattcagaataatttcaataaactgattatatoscgctggtac 405
Db 174 attattatgcgaagttaaattcagaataatttcaataaactgattatatoscgctggtac 233
QY 406 attgcgctag 415
Db 234 attgcgctag 243

RESULT 9
US-08-817-188-5
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTriB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' q7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: pSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3059)..(5056)
; OTHER INFORMATION: uidA: region coding for beta-glucuronidase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4483)..(4671)
; OTHER INFORMATION: IV2: region corresponding to the second intron of
; OTHER INFORMATION: the ST-LS1 gene
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (5067)..(5502)
; OTHER INFORMATION: P35S: 35S promoter region of CamV
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5533)..(5560)
; OTHER INFORMATION: LB: left border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTriB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5058)..(5059)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5077)..(5078)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5476)..(5479)
; OTHER INFORMATION: region with unknown sequence (may contain up to 20
; OTHER INFORMATION: nucleotides)
; OTHER INFORMATION: nucleotides)
US-08-817-188-5

```

```

Query Match 45.4%; Score 188.4; DB 3; Length 5560;
Best Local Similarity 99.5%; Pred. No. 1.4e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 226 gtacatggccgataagaagaagcaatttgatagatgttaattcccatcttgaagaataat 285
Db 41 gtacatggctgataagaagaagcaatttgatagatgttaattcccatcttgaagaataat 100
QY 286 agtttaaatattattgataaaataacaagtcagggtattattatagtcacgaacaaacataa 345
Db 101 agtttaaatattattgataaaataacaagtcagggtattattatagtcacgaacaaacataa 160
QY 346 attattatgcgaagttaaattcagaataatttcaataaactgattatatoscgctggtac 405
Db 161 attattatgcgaagttaaattcagaataatttcaataaactgattatatoscgctggtac 220
QY 406 attgcgctag 415
Db 221 attgcgctag 230

```

```

RESULT 10
US-08-232-016-23/c
; Sequence 23, Application US/08232016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016

```

```

; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91402920.2
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pPS0212
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated modified bt2 (cryIab) gene, also designated as the
; OTHER INFORMATION: cryIab6 gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1793..2026
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2396..2921
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2922..3581
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3582..4407
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5600..6457
; OTHER INFORMATION: /note= "Sequence complementary to
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; Patent No. 5952547
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7071..7566
; OTHER INFORMATION: /note= "TR1' and TR2 promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader with
; OTHER INFORMATION: respect to sequence of pJD884 of SEQ ID NO. 22."
; US-08-232-016-23
;
Query Match 45.4%; Score 188.4; DB 2; Length 7566;
Best Local Similarity 99.5%; Pred. No. 1.5e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 226 gtacatgccgataagaagcaattgttagatgttaattcccatcttgaagaataat 285
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

Db 2070 GTACATGGTCGATAAGAAAGCAATTTGTAGATGTTAAATTCCTCTTTGAAGAAATAT 2011
QY 286 agtttaaatattattgataaaatacaagtcaggtattattatagtcacagcaaaacataa 345
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2010 AGTTTAAATATTATTGATAAAATACAAAGTCAGGTATTATTATAGTCCAAACAAAACATAA 1951
QY 346 atttattgatgcaagtttaattcagaaatatttcaataactgattatataatcagctggtag 405
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1950 ATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATATCAGCTGGTAC 1891
QY 406 attgccgtag 415
||||||| |||||||
Db 1890 ATTGCCGTAG 1881

RESULT 11
US-08-232-016-22/c
; Sequence 22, Application US/082332016
; Patent No. 5952547
; GENERAL INFORMATION:
; APPLICANT: CORNELISSEN, Marc
; APPLICANT: SOETAERT, Piet
; APPLICANT: STAM, Maïke
; APPLICANT: DOCKX, Jan
; APPLICANT: VAN AARSEN, Roel
; TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
; TITLE OF INVENTION: PLANT CELLS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,016
; FILING DATE: 03-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91402920.2
; FILING DATE: 30-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 92400820.4
; FILING DATE: 25-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K
; REGISTRATION NUMBER: P39,300
; REFERENCE/DOCKET NUMBER: 010830-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid DNA designated as pJD884
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1869
; OTHER INFORMATION: /note= "Coding region of a
; OTHER INFORMATION: truncated bt2 (cryIab) gene, also designated as the bt884
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: 1877..2110
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA gene 7."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2480..3005
; OTHER INFORMATION: /note= "35S promoter sequence
; OTHER INFORMATION: derived from Cauliflower mosaic virus."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3006..3665
; OTHER INFORMATION: /note= "Coding sequence of
; OTHER INFORMATION: chloramphenicol acetyl transferase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3666..4491
; OTHER INFORMATION: /note= "3' regulatory sequence
; OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteriu
; OTHER INFORMATION: T-DNA octopine synthase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5684..6541
; OTHER INFORMATION: /note= "Sequence complementary to
; Patent No. 5952547
; OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7155..7639
; OTHER INFORMATION: /note= "TR1' and TR2' promoter
; OTHER INFORMATION: derived from Agrobacterium T-DNA."
;
US-08-232-016-22

```

```

Query Match      45.4%; Score 188.4; DB 2; Length 7639;
Best Local Similarity 99.5%; Pred. No. 1.5e-36;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 226 gtacatggccgataagaagaaaggcaattttagatgttaattcccatcttgaagaataat 285
Db 2154 GTACATGTCGATGAAGAAAGGCAATTTGTAGATGTTAATCCCATCTTGAAGAATAAT 2095

Qy 286 agtttaaatattattgataaaatacaacgacggtattatagtcacgaagcaaaacataa 345
Db 2094 AGTTTAAATATTATTGATATAAAATAACAGTCAGGTATTATAGTCCCAAGCAAAACATAA 2035

Qy 346 atttatgtgcaggttttaattcagaataattttcaataactgattataccagctggtac 405
Db 2034 ATTATTGATGCAAGTTTAAATTACAGAAATATTTCATAACTGATTATATACAGCTGTAC 1975

Qy 406 attgcccgtag 415
Db 1974 ATTGCCGTAG 1965

```

```

RESULT 12
US-08-894-440-4
; Sequence 4, Application US/08894440
; Patent No. 6025546
; GENERAL INFORMATION:
; APPLICANT: PLANT GENETIC SYSTEMS N.V.
; TITLE OF INVENTION: Method to obtain male sterile plants
; FILE REFERENCE: NNSCOR
; CURRENT APPLICATION NUMBER: US/08/894,440
; CURRENT FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 5864
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of

```

```

; OTHER INFORMATION: plasmid pTCO113
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((1)..(25))
; OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((98)..(330))
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((331)..(882))
; OTHER INFORMATION: region coding for phosphinothricin acetyl
; OTHER INFORMATION: transferase (bar)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((883)..(2608))
; OTHER INFORMATION: promoter of small subunit gene of Rubisco of
; OTHER INFORMATION: Arabidopsis (Pssu)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((2659)..(3031))
; OTHER INFORMATION: region containing polyadenylation signal of
; OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
; OTHER INFORMATION: (3'nos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3032)..(3367))
; OTHER INFORMATION: region coding for barnase of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((3368)..(4877))
; OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum (PTA29)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4924)..(5216)
; OTHER INFORMATION: promoter of nopaline synthase gene of
; OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5217)..(5489)
; OTHER INFORMATION: region coding for barstar of Bacillus
; OTHER INFORMATION: amyloliquefaciens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (5490)..(5765)
; OTHER INFORMATION: region containing polyadenylation signal of gene 7
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: Complement((5840)..(5864))
; OTHER INFORMATION: left border of Agrobacterium T-DNA
;
US-08-894-440-4

```

```

Query Match      45.0%; Score 186.8; DB 3; Length 5864;
Best Local Similarity 98.9%; Pred. No. 3.4e-36;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 226 gtacatggccgataagaagaaaggcaattttagatgttaattcccatcttgaagaataat 285
Db 54 gtacatggctcgataagaagaaaggcaattttagatgttaattcccatcttgaagaataat 113

Qy 286 agtttaaatattattgataaaatacaacgacggtattatagtcacgaagcaaaacataa 345
Db 114 agtttaaatattattgataaaatacaacgacggtattatagtcacgaagcaaaacataa 173

Qy 346 atttatgtgcaggttttaattcagaataattttcaataactgattataccagctggtac 405
Db 174 atttatgtgcaggttttaattcagaataattttcaataactgattataccagctggtac 233

```

QY 406 attgcgctag 415  
|||||  
Db 234 attgcgctag 243

RESULT 13  
US-08-453-104-22/c  
; Sequence 22, Application US/08453104  
; Patent No. 5633446  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453.104  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-453-104-22

Query Match 42.7%; Score 177; DB 1; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 239 aagaaaggcaattttagatgttaattcccatcttgaagaaataatagtttaaatattt 298  
|||||  
Db 3201 AGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAAATATAGTTAAATATTT 3142  
QY 299 attgataaataacaagtcaggtattatagtcacagcaaaaacataaaatttattgatgca 358  
|||||  
Db 3141 ATTGATAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACATAAATTTATTGATGCA 3082

QY 359 agtttaaatcagaaatatttcaataactgattatatactagctggtacattgcgctag 415  
|||||  
Db 3081 AGTTTAAATTCAGAAATATTTCATAACTGATTATATATCATCAGCTGGTACATTGCCGTAG 3025

RESULT 14  
US-08-694-824-22/c  
; Sequence 22, Application US/08694824  
; Patent No. 5877306  
; GENERAL INFORMATION:  
; APPLICANT: CORNELISSEN, Marc  
; APPLICANT: SOETAERT, Piet  
; APPLICANT: STAM, Maïke  
; APPLICANT: DOCKX, Jan  
; TITLE OF INVENTION: MODIFIED BACILLUS THURINGIENSIS  
; TITLE OF INVENTION: INSECTICIDAL - CRYSTAL PROTEIN GENES AND THEIR EXPRESSION  
; TITLE OF INVENTION: IN PLANT CELLS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: George Mason Bldg., Washington & Prince Sts.  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/694,824  
; FILING DATE: 09-AUG-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/937,869  
; FILING DATE: 16-DEC-1992  
; APPLICATION NUMBER: GB 90401055.0  
; FILING DATE: 18-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa S  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 010830-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 2151..2155  
; OTHER INFORMATION: /note= "Nucleotides 2151-2155  
; OTHER INFORMATION: wherein N is not known."  
US-08-694-824-22

Query Match 42.7%; Score 177; DB 2; Length 3201;  
Best Local Similarity 100.0%; Pred. No. 6.5e-34;  
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 239 aagaaaggcaattttagatgttaattcccatcttgaagaaataatagtttaaatattt 298  
|||||  
Db 3201 AGAAAGGCAATTGTAGATGTTAATCCCATCTTGAAGAAATATAGTTAAATATTT 3142  
QY 299 attgataaataacaagtcaggtattatagtcacagcaaaaacataaaatttattgatgca 358  
|||||  
Db 3141 ATTGATAAATAACAAGTCAGGTATTATAGTCCAAAGCAAAACATAAATTTATTGATGCA 3082







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 12:03:22 ; Search time 1397.28 Seconds  
(without alignments)  
3191.557 Million cell updates/sec

Title: US-09-698-903B-8  
Perfect score: 415  
Sequence: 1 gtcgagttgtgttcattga.....cagctggtacatgcgtag 415

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	em_estfun:*
2:	em_esthum:*
3:	em_estim:*
4:	em_estom:*
5:	em_estpl:*
6:	em_estba:*
7:	em_estro:*
8:	em_estov:*
9:	em_htc:*
10:	gb_est1:*
11:	gb_est2:*
12:	gb_htc:*
13:	gb_gss:*
14:	em_gss_fun:*
15:	em_gss_hum:*
16:	em_gss_inv:*
17:	em_gss_pln:*
18:	em_gss_pro:*
19:	em_gss_rod:*
20:	em_gss_vrt:*
21:	em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB	ID Description
1	63	15.2	734 13 CNS010MP AL099163 Drosophil
2	61	14.7	1101 13 CNS0042W AL055440 Drosophil
3	57.2	13.8	844 13 CNS03D0I AL238491 Tetraodon
4	57.2	13.8	1101 13 CNS000B8 AL063632 Drosophil
5	57.2	13.8	1101 13 CNS00238 AL097166 Drosophil
6	55.4	13.3	1101 13 CNS016LI AL106896 Drosophil
7	55	13.3	537 13 AQ506817 RPCI-11-2
8	54.8	13.2	987 13 CNS014PQ AL104456 Drosophil
9	54.8	13.2	1101 13 CNS003BD AL064091 Drosophil
10	54.6	13.2	980 13 CNS00JG1 AL076232 Drosophil
11	54.6	13.2	1101 13 CNS00EVL AL069706 Drosophil
12	54.4	13.1	893 13 CNS013XE AL103436 Drosophil

13	54	13.0	905	13	CNS00KHx	AL077798 Drosophil
14	53.4	12.9	500	10	AU087444	AU087444
15	53.4	12.9	970	13	CNS0182A	AL108796 Drosophil
16	53.2	12.8	1001	13	CNS0155H	AL105023 Drosophil
17	53.2	12.8	1101	13	CNS0106X	AL098595 Drosophil
18	53	12.8	1101	13	CNS016LI	AL106896 Drosophil
19	52.6	12.7	734	13	CNS010MP	AL099163 Drosophil
20	52.6	12.7	1013	13	CNS00J7I	AL075824 Drosophil
21	52.6	12.7	1101	13	CNS00EPO	AL069493 Drosophil
22	52.4	12.6	524	13	CNS01U90	AL167541 Tetraodon
23	52.4	12.6	996	13	CNS00FUH	AL071063 Drosophil
24	52.2	12.6	639	13	CNS038CX	AL232458 Tetraodon
25	51.8	12.5	928	13	CNS00DKY	AL071865 Drosophil
26	51.8	12.5	1101	13	CNS0039G	AL063921 Drosophil
27	51.8	12.5	1101	13	CNS00FMC	AL070972 Drosophil
28	51.6	12.4	992	13	CNS0562R	AL322812 Tetraodon
29	51.6	12.4	1043	13	CNS0145P	AL103735 Drosophil
30	51.6	12.4	1101	13	CNS00EO7	AL069440 Drosophil
31	51.4	12.4	876	13	CNS009G1	AL053529 Drosophil
32	51.2	12.3	678	13	CNS02A0C	AL187941 Tetraodon
33	51.2	12.3	1001	13	CNS01400	AL103554 Drosophil
34	51.2	12.3	1101	13	CNS00EVL	AL069706 Drosophil
35	50.6	12.2	854	11	BF274512	BF274512 GA_EB002
36	50.6	12.2	963	10	AL566565	AL566565
37	50.6	12.2	1101	13	CNS00CYH	AL060100 Drosophil
38	50.6	12.2	1101	13	CNS00Z2U	AL097152 Drosophil
39	50.4	12.1	728	13	AQ272964	AQ272964 nbxb0028P
40	50.4	12.1	1101	13	CNS003BB	AL064089 Drosophil
41	50.4	12.1	1200	13	CNS016CO	AL106578 Drosophil
42	50.2	12.1	681	13	B60190	B60190 CIT-HSP-200
43	50	12.0	563	13	AQ326762	AQ326762 nbxb0038D
44	50	12.0	788	13	BH126604	BH126604 BARC-Satt
45	50	12.0	987	13	CNS014PQ	AL104456 Drosophil

ALIGNMENTS

RESULT 1  
LOCUS CNS010MP 734 bp DNA GSS 26-JUL-1999  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL099163  
VERSION AL099163.1 GI:5610774  
KEYWORDS GSS:  
SOURCE fruit fly.  
ORGANISM Plasmid Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 734)  
AUTHORS Genoscope.  
TITILE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.  
FEATURES  
source Location/Qualifiers  
1..734  
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN04L20"

BASE COUNT	288 a	62 c	2 g	211 t	171 others
ORIGIN	/note="end : T7"				

Query Match 15.2%; Score 63; DB 13; Length 734;  
Best Local Similarity 41.1%; Pred. NO. 0.0079;  
Matches 85; Conservative 43; Mismatches 79; Indels 0; Gaps 0;

QY	189	aagtattataataattataataataatgatgtaggtacatgcgcgataagaaaagc	248
Db	53	AMATATGTATATATATATATAWAATWATAWAAAAMAMWTTCAAATTTMMAAAARAAWAAAW	112
QY	249	aattgtagatgtaattcccatcttgaagaataagtttaaatattattattgataaaa	308
Db	113	WTAAHMAWAATATWATTTTNTTTTHMAAWATATTTTTTTTATTTTAAATTTTTTTTAA	172
QY	309	taacaagtcagggtattatagtcgaagcaaacataaaattttatgcaagtttaaat	368
Db	173	AWMMMTYAWATTTTWTWATWAAAMWTTWAAAWAAWAAWAAWATTTATATATWTTTAAWAA	232
QY	369	cagaataattccaataactgattatat	395
Db	233	AWAAATATWMAAAAAAMWWWWAT	259

[illegible]

ACCESSION AL055440  
VERSION AL055440.1 GI:4932241  
KEYWORDS GSS

SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 1101) Genoscope.

**TITLE** Direct Submission  
**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial *EcoRI* digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

source
1. 1101
/organism="Drosophila melanogaster"
/bb_ref="taxon/7227"
/clone_lib="RPCI-98"
/clone="BACR11E08"

```

BASE COUNT	/note="end : T7"		
ORIGIN	294 a	74 c	99 g
	381 t	253 others	

Query Match 14.7%; Score 61; DB 13; Length 1101;

Best Local Similarity 38.8%; Pred. No. 0.018;  
Matches 149; Conservative 51; Mismatches 183; Indels 1; Gaps 1;

Qy	14	ttcatgatgtttggggttgactcttccaccattacatatatgaaactcttcaogtagaac	73
Db	1078	WTWTWTWTATTTTAAATTTTAAATTTTATATATATWAAAWATTTTAAWATTTTWTAAAC	101
Qy	74	aactcacaagcattaatcatgtcttcataataaatatgtacattatcatgtatatatacaog	133
Db	1018	AAAAAYAAATATTTTTTTTAAWTAAATATATAAATWTWAAATTTAAWTTTATWATATAWA	959
Qy	134	tatacaaatagtagcgagaagaatcccatgtaaacgacgaggggcaccatgggttccaagta	193
Db	958	TATWAAATAAWATHTWTAATAATAAATAAAWATAAAWATATATWTAATAAAWATTTTAAWTTATTATTK	899
Qy	194	ttatataaattat-aatatataattatgtgtaggatgtacatggccgataagaaagcgcaatt	252
Db	898	ATWTAATATTTAAAAATTTATATTTWATGAATAAAATTTAAATAATATAANTATATATTTKAT	839
Qy	253	tgtagatgttaatcccatcttgaagaataatagttttaaattattatgtataaaataaac	312
Db	838	TTAAWATATATBTTSATATATAATAAAATATAAAATAAANATAAAMWAAWATAANWAT	779
Qy	313	aagtcgaggtattatagtcaccaagcaaaacataaaattattgtatgcagttttaaattcaga	372
Db	778	ATATATATWATWAAATTTTAAWATWTAANWATTKAWATATAAWATTTTAAWTTTAAWTTTANA	719
Qy	373	aatattccaatcaactgattatc	396
Db	718	AATMTAAWAMTGTATATATATTTTH	695

RESULT 3  
CNS03D0I/C

LOCUS	CNS03D01	844 bp	DNA	GSS	15-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 015114 of library G from Tetraodon nigroviridis, genomic survey sequence.				

sequence.  
ACCESSION AL238491  
VERSION AL238491.1 GI:7997626  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis.

Tetrapodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.

REFERENCE  
AUTHORS  
1 (bases 1 to 844)  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,  
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
Weissenbach, J.

# Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL  
REFERENCE  
Unpublished  
2 (bases 1 to 844)

**AUTHORS**  
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
Saurin, W. and Weissenbach, J.

**TITLE** Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL  
REFERENCE  
AUTHORS  
Unpublished  
3 (bases 1 to 844)  
Genoscope.

REMARKS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBDJ databases
COMMENT	This sequence is a single read and was generated on a

sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

## FEATURES

```
1. 044
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
```

```
/clone="015L14"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG015DF07SP1-end :
PUC-Orig"
BASE COUNT      152 a   79 c   89 g   484 t   40 others
ORIGIN

Query Match      13.8%; Score 57.2; DB 13; Length 844;
Best Local Similarity 47.0%; Pred. No. 0.099;
Matches 155; Conservative 6; Mismatches 169; Indels 0; Gaps 0;

Qy 56 actcttagcgatgagacacactcacacattatcatctgttcataataatatgtacat 115
Db 487 AGTGTGGCATAATAACAAGTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 428
Qy 116 tatcgttatatacacgttatatacaatagtagcgaagaatccatgtaaacgacgggg 175
Db 427 AATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 368
Qy 176 gcaccatggttccaagtattataattataattataattatggttaggtgacatggcc 235
Db 367 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 308
Qy 236 gataagaagaagcaattttagatgttaattcccatcttgaaagaataatgtttaata 295
Db 307 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 248
Qy 296 tttattgataaataacaagtcaggattattatagtcgaagcaaaacataattttgat 355
Db 247 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 188
Qy 356 gcaagtttaaatcagaataatttcaataa 385
Db 187 TCTAAAAAATAATCATAAACATTAAATCA 158

RESULT 4
CNS000B8/c 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR01A24 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063632
VERSION AL063632.1 GI:4938680
KEYWORDS fruit fly,
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
Location/Qualifiers
1..1101
source
```

```
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR01A24"
/note="end : TET3"
BASE COUNT      228 a   114 c   110 g   512 t   137 others
ORIGIN

Query Match      13.8%; Score 57.2; DB 13; Length 1101;
Best Local Similarity 39.5%; Pred. No. 0.094;
Matches 147; Conservative 56; Mismatches 165; Indels 4; Gaps 2;

Qy 43 attacattgaaactcttcacggatgagacacactcacacattatcatgttcataata 102
Db 1087 AWAANAATAATTAATATTTWAAAAAATAAATAAATAAATAAATAAATAAATAA 1029
Qy 103 aatatgtacattatcgttatatacacgttatatacaatagtagcgaagaataccatgt 162
Db 1028 AAAAAATAAAATTTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 969
Qy 163 aaagcagcagggggccaccatggttccaagtattataattataattataattatggtag 222
Db 968 ATWTAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 909
Qy 223 gatgtacatggccgataagaagaaggcaattttagatgttaattcccatcttgaagaaa 282
Db 908 TTTAAWATWAAAAATWAAAAATTTAAWTTAAWTTAAWTTAAWTTAAWTTAAWTTAA 849
Qy 283 tatagtttaaat---attattgataaaatacaacagtcaggtattatagtcgaagcaaaa 339
Db 848 AATWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 789
Qy 340 acataaaatttattgatgcaagtttaattcagaataatttcataactgattatcagc 399
Db 788 ATWAAATTTACMAAATYWAATTTAAATTAATAAATAAATAAATAAATAAATAAATAA 729
Qy 400 tggtagatggc 411
Db 728 AATTAAWTTTCM 717

RESULT 5
CNS00Z38/c 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN01A24 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL097166
VERSION AL097166.1 GI:5608777
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmod Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (BDGP) -
http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
Location/Qualifiers
1..1101
source
```

```

/organism="Drosophila melanogaster"
/plasmid="pBalORXc11"
/db xref="tspan 722"
/clone.lib="pcBAC"
/clone="BACN0134"
/note="end : SP6"

308 a 152 c 162 g 386 t 93 others
BASE COUNT
ORIGIN

```

Query Match 13.8%; Score 57.2; DB 13; Length 1101;  
Best Local Similarity 41.4%; Pred. NO. 0.094;  
Matches 87; Conservative 38; Mismatches 85; Indels 0; Gaps 0;

[illegible]

Qy 366 attcagaaatatttcaataactgattatat 395  
          | : | : | : | : | : | : | :  
Db 858 WTWTWTWTWTTTTMMMHMMHHHAYTAWAWT 829

RESULT	6
CNS016LI/c	
LOCUS	CNS016LI 1101 bp DNA GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BAC16022 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL106896  
VERSION AL106896.1 GI:5624374

KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Plasamid <i>Drosophila melanogaster</i> <i>Eukaryota</i> ; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; <i>Pterygota</i> ; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; <i>Drosophila</i> . 1 (bases 1 to 1101)
REFERENCE	

AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:seqrefgenoscope.cns.fr">seqrefgenoscope.cns.fr</a> - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <a href="http://www.edgp.ebi.ac.uk">http://www.edgp.ebi.ac.uk</a> -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES	source	Location/Qualifiers
BASE COUNT	203 a	220 c 84 g 158 t 436 others
ORIGIN		1..1101
		/organism="Drosophila melanogaster"
		/protein="pBelobAC11"
		/db_xref="taxon:71217"
		/clone_lib="DrosBAC"
		/clone="BACN1602"
		/notes="end : 177"

Query Match 13.3%; Score 55.4; DB 13; Length 1101;  
Best Local Similarity 30.7%; Pred. No. 0.21;

Matches	96;	Conservative	79;	Mismatches	138;	Indels	0;	Gaps	0;
QY	43	attacatatgaaactcttagcgatgagacaacaccagaacattaatcaatgcgttcataata	102						
Db	1050	ANWAAAWAANAATTTAYATTHTTAATHANTWTWCATAHWWWWTTTTTATWAATH	991						
QY	103	aatatgtcattcacgttatatacacgctatcacaaatgatgcgaagaatccactgt	162						
Db	990	TATTWATAAHWTATYTWTWHNATYWTWTYTATWAAHAANAATASTPAAIYWIAHTATW	931						
QY	163	aaagcagcaggggccaccatggtttccaagtattataataattaataattatggtag	222						
Db	930	AWANATAATTWTWTWTAAWTAAATATWATWAAANTAAATWWWAAHWAAWTAYAWW	871						
QY	223	gatgtcacatggcccataagaagaaggcaatttgtagtggtaattccccactcttgagaaa	282						
Db	870	WWNAWTATWNTATWATAATAAWATATWTAATTTATATWAAAAATATATATWAAAAAAA	811						
QY	283	tatgatttaaataatttatgtataaatacaaacgcaggtcaggtattatagcccaagcga	342						
Db	810	THAAWAAWAAATATWAAAAAAAHHMTAAAWATAAAAVAAAAWAAAAWAAAAAAH	751						
QY	343	taaatatttgat	355						
Db	750	ATMAATTTTWT	738						

RESULT	7
AQ506817/c	
LOCUS	AQ506817 537 bp DNA GSS 29-APR-1999
DEFINITION	RPCI-11-281J17.TV RPCI-11 Homo sapiens genomic clone RPCI-11-281J17 , DNA sequence.

ACCESSION	AQ0506817		
VERSION	AQ0506817.1	GI:4711564	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Karyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.		
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building		
JOURNAL	Unpublished (1997)		
COMMENT	Other_GSSs: RPCI-11-281J17.TJ Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbeatigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACAP Resources ( <a href="http://bacpac.med.buffalo.edu/ordering">http://bacpac.med.buffalo.edu/ordering</a> ) or from Research Genet cs ( <a href="mailto:info@resgen.com">info@resgen.com</a> ). BAC end search page: <a href="http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html">http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html</a> . seq primer: TJ		

FEATURES	source	Class: BAC ends.
	Location/Qualifiers	
	1. .537	
	/organism="Homo sapiens"	
	/db_xref="GDB:760752"	
	/db_xref="taxon:9606"	
	/clone="RPCI-11-281J17"	
	/clone_lib="RPCI-11"	
	/sex="Male"	
	/cell_type="Lymphocytes"	
	/note="Vector: pRACE3.6; Site_1: ECORI;	
	RPCI11 Human Male BAC Library"	
BASE COUNT	216 a 11 c 30 g 280 t	



[illegible]



BASE COUNT	419 a	91 c	60 g	299 t	232 others
ORIGIN					
Query Match	13.2%; Score 54.6; DB 13; Length 1101;				
Best Local Similarity	33.7%; Pred. No. 0.29; Mismatches 156; Indels 0; Gaps 0;				
Matches lll; Conservative	62;				
Qy 69	agaaacaactcaagcattatcatgtttccataataatgatgacattacgtatatat 128         : : :         : : :         : : :         : : :				
Db 731	AWAATAAAWAAATWAATAATATTTTATATATATTTTAAWWATWWWATATATAT 790         : : :         : : :         : : :         : : :				
Qy 129	aacogtatacaaatagtgcgaagaataccatgtaaacgaggcgccaccatgttc 188     : : :     : : :         : : :         : : :         : : :				
Db 791	AWAATAAAWAAWATAAATAATATATATATTAATAAATAAATAAATAAATAA 850     : : :     : : :         : : :         : : :         : : :				
Qy 189	aagtattataataataattatgtagtgtaggtacatggccgataagaagaagc 248         : : :         : : :         : : :         : : :				
Db 851	AAAAATTGAATATWAATATATTAATAAATAAATAAATAAATAAATAAATAA 910         : : :         : : :         : : :         : : :				
Qy 249	aattttagatgtcaattccccctcttgaagaataatagtttaaattatttgtataaa 308     : : :     : : :         : : :         : : :         : : :				
Db 911	TAAWAAWAAAATAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAA 970         : : :         : : :         : : :         : : :				
Qy 309	taacaagtcaggtattatagtcgaagcaaaaacataaatttatgtgcaagtttaaa 368     : : :     : : :         : : :         : : :         : : :				
Db 971	WANTWTATATATTTATATATATATATATATATATATATATATATATATAT 1030         : : :         : : :         : : :         : : :				
Qy 369	cagaataattccaataacgtattatca 397         : : :         : : :         : : :         : : :				
Db 1031	WTAAT 1059         : : :         : : :         : : :         : : :				
RESULT 12					
CNS013XE	CNS013XE 893 bp DNA GSS 26-JUL-1999				
LOCUS	Drosophila melanogaster genome survey sequence SP6 end of BAC				
DEFINITION	AL1010102 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL103436				
VERSION	AL103436.1 GI:5615047				
KEYWORDS	GSS.				
SOURCE	fruit fly.				
ORGANISM	Plasmod Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
TITLE	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
JOURNAL	1 (bases 1 to 893)				
COMMENT	Genoscope. Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr) - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pbeloBAC11.				
FEATURES	Location/Qualifiers				
source	1..893 /organism="Drosophila melanogaster" /plasmid="pbeloBAC11" /db_xref="taxon:7227" /clone_lib="DrosBAC" /clone="BACN10102" /note="end : SP6"				
BASE COUNT	372 a	21 c	33 g	308 t	159 others
ORIGIN					



```
Db 75 AAAAAAAAAACCCCTAATAATAAACNACAAAAATTTAAAAATAATAATAAAAAATAAAA 134
Qy 159 atgtaaagcagcaggggcaccatggtttcaagtattataataattataattataattatg 218
   ||:|||||:| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 ATAWAAAAAAATWTTTATWAAWAAAAATATAWATATWAAAAAAANTTNCTAATTTT 194
Qy 219 gtaggatgtacatggccgataagaaggaagcattttagtagatgttaattcccatcttgaaa 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 ATTATNATAAAATAAANCTTATAAAAAAAATAAAAAATAAAAAATAAAAAACRATATAAAAA 254
Qy 279 gaaatagtttaaatatttattgataaaataacaagtcagggtattatagtcacaagcaaa 338
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 ATAAAAAAWATAAAAAAAATAAWAAAAAAWAAAAAAWAAAAATAAAAAATAAAAAATA 314
Qy 339 aacataaatttattgatgcaagtttaaatcagaata 376
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 AAAAWTAAAAATATAAAAAAAATTAATWAAWAAAAAAA 352
```

Search completed: February 25, 2002, 12:26:59  
Job time: 1417 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:00:20 ; Search time 2331.3 Seconds  
(without alignments)  
169.833 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24  
Sequence: 1 tcagaagtatcagcagacctccacc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_om.\*
- 20: em\_or.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_ro.\*
- 26: em\_sts.\*
- 27: em\_sy.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htgo\_hum.\*
- 31: em\_htgo\_inv.\*
- 32: em\_htgo\_rod.\*
- 33: em\_htg\_hum.\*
- 34: em\_htg\_inv.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24	100.0	24	6	AX127756	AX127756 Sequence 1
2	24	100.0	270	6	A71431	A71431 Sequence 1
3	24	100.0	323	6	A87282	A87282 Sequence 7
4	24	100.0	340	6	A21284	A21284 Artificial
5	24	100.0	474	1	BABARSTA	X15545 Bacillus am
6	24	100.0	4032	6	A71435	A71435 Sequence 5
c 7	24	100.0	4808	6	AR007527	AR007527 Sequence
c 8	24	100.0	4808	6	AR084093	AR084093 Sequence
c 9	24	100.0	4832	6	AX172441	AX172441 Sequence
10	24	100.0	5865	6	AX127748	AX127748 Sequence
c 11	24	100.0	6555	6	AR007512	AR007512 Sequence
c 12	24	100.0	6555	6	AR084078	AR084078 Sequence 3
13	20.8	86.7	273	6	A71433	A71433 Sequence 6
14	20.8	86.7	563	6	A71436	A71436 Sequence 7
15	20.8	86.7	5349	6	A71437	A71437 Sequence 10
16	20.8	86.7	5611	6	A71440	A71440 Sequence 10
c 17	19.8	82.5	130235	8	AC008007	AC008007 Genomic s
c 18	18.2	75.8	426	9	AF062270	AF062270 Homo sapi
19	18.2	75.8	74213	2	AC025320	AC025320 Homo sapi
c 20	18.2	75.8	124990	9	HS434P1	Z97056 Human DNA s
21	18.2	75.8	165471	2	AC015713	AC015713 Homo sapi
22	18.2	75.8	234131	2	AC093357	AC093357 Mus muscu
23	17.8	74.2	1905	1	COXCYTSTN	M36338 C.burnetii
c 24	17.8	74.2	173846	2	AC087781	AC087781 Mus muscu
25	17.8	74.2	238737	2	AC084073	AC084073 Mus muscu
c 26	17.6	73.3	731	9	HSDCTN08	AF086933 Homo sapi
c 27	17.6	73.3	3534	6	AX015400	AX015400 Sequence
c 28	17.6	73.3	4153	9	HSDYNACTN	X98801 H.sapiens m
c 29	17.6	73.3	12426	9	HSDCNTA13	AF064205 Homo sapi
c 30	17.6	73.3	12956	1	AE006198	AE006198 Pasteurel
c 31	17.6	73.3	13223	1	AVU49859	U49859 Anabaena va
c 32	17.6	73.3	34646	3	U00066	U00066 Caenorhabdi
c 33	17.6	73.3	46843	8	SPBC530	AL023634 S.pombe c
c 34	17.6	73.3	50502	9	HS4494016	AL117328 Human DNA
35	17.6	73.3	52358	9	HS27C10	AL031803 Human DNA
c 36	17.6	73.3	69506	8	AC012680	AC012680 Arabidops
c 37	17.6	73.3	95769	8	AC013430	AC013430 Genomic s
c 38	17.6	73.3	129538	9	AC008854	AC008854 Homo sapi
39	17.6	73.3	139152	8	AP002525	AP002525 Oryza sat
40	17.6	73.3	141198	2	AC023898	AC023898 Mus muscu
c 41	17.6	73.3	147722	2	AC016783	AC016783 Homo sapi
c 42	17.6	73.3	157993	2	AC073354	AC073354 Homo sapi
c 43	17.6	73.3	167263	2	AC063966	AC063966 Homo sapi
c 44	17.6	73.3	167943	2	AC026322	AC026322 Homo sapi
c 45	17.6	73.3	170081	2	AC055714	AC055714 Homo sapi

ALIGNMENTS

RESULT 1	AX127756	24 bp	DNA	PAT	15-MAY-2001
LOCUS	AX127756	Sequence 9 from Patent WO0131042.			
DEFINITION	AX127756				
ACCESSION	AX127756.1	GI:14134403			
VERSION					
KEYWORDS		synthetic construct.			
SOURCE		synthetic construct			
ORGANISM		artificial sequence.			
REFERENCE		1 (bases 1 to 24)			
AUTHORS		Weston,B. and de Beuckeleer,M.			
TITLE		Male-sterile brassica plants and methods for producing same			
JOURNAL		Patent: WO 0131042-A 9 03-MAY-2001;			
FEATURES		Aventis CropScience N.V. (BE)			
source		Location/Qualifiers			
		1..24			
		/organism="synthetic construct"			
		/db_xref="taxon:32630"			
		/note="primer MDB8"			

BASE COUNT 7 a 9 c 4 g 4 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 24;  
Best Local Similarity 100.0%; Pred. No. 0.086;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 1 TCAGAAGTATCAGCGACCTCCACC 24

## RESULT 2

A71431 A71431 270 bp DNA PAT 07-MAY-1999  
LOCUS  
DEFINITION Sequence 1 from Patent WO9810081.  
ACCESSION A71431  
VERSION A71431.1  
KEYWORDS GI:4775044

SOURCE  
ORGANISM Bacillus amyloliquefaciens.

REFERENCE  
AUTHORS Bacteria; Firmicutes; Bacillus/Clostridium group;  
TITLE Bacillus/Staphylococcus group; Bacillus.

JOURNAL  
Micheels, F. and Williams, M.  
TITLE IMPROVED BARSTAR GENE  
PATENT: WO 9810081-A 1 12-MAR-1998;  
MICHIELS FRANK (BE)

FEATURES  
source  
1..270  
Location/Qualifiers

/organism="Bacillus amyloliquefaciens"  
/db\_xref="taxon:1390"

1..>270  
/function="INHIBITOR OF BARNASE"

/codon\_start=1  
/transl\_table=1  
/product="BARSTAR"

/protein\_id="CAB42577.1"

/db\_xref="GI:4775045"

/translation="MKKAVINGQIRISDLHTLKKELALPEYYGENLDALWDCLTG  
WVEYPLVLEWRQEQSKLTENGAESVLQVFRKAEGCDITILS"

BASE COUNT 80 a 58 c 74 g 58 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 270;  
Best Local Similarity 100.0%; Pred. No. 0.095;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 32 TCAGAAGTATCAGCGACCTCCACC 55

## RESULT 3

A87282 A87282 323 bp DNA PAT 22-JAN-2000  
LOCUS  
DEFINITION Sequence 7 from Patent WO9837211.  
ACCESSION A87282  
VERSION A87282.1  
KEYWORDS GI:6736047

SOURCE  
ORGANISM unidentified.

REFERENCE  
1 (bases 1 to 323)

AUTHORS Huttner, E. and Betzner, A.S.

TITLE PROTEIN COMPLEMENTATION IN TRANSGENIC PLANTS

JOURNAL Patent: WO 9837211-A 7 27-AUG-1998;

GENE SHEARS PTY LTD (AU); HUTTNER ERIC (AU)

FEATURES  
source  
1..323  
Location/Qualifiers

/organism="unidentified"

## CDS

/db\_xref="taxon:32644"  
1..273  
/codon\_start=1  
/product="BARSTAR"  
/protein\_id="CAB69371.1"  
/db\_xref="GI:6736048"

/translation="MKKAVINGQIRISDLHTLKKELALPEYYGENLDALWDCLTG  
WVEYPLVLEWRQEQSKLTENGAESVLQVFRKAEGCDITILS"

BASE COUNT 98 a 69 c 87 g 69 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 323;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 32 TCAGAAGTATCAGCGACCTCCACC 55

## RESULT 4

A21284 A21284 340 bp DNA PAT 31-MAY-1994  
LOCUS  
DEFINITION Artificial barstar gene.  
ACCESSION A21284  
VERSION A21284.1  
KEYWORDS GI:514151

SOURCE  
ORGANISM unidentified.

REFERENCE  
1 (bases 1 to 340)

AUTHORS Mariani, C., Leemans, J. and De Greef, W.

TITLE Plants with modified flowers

JOURNAL Patent: EP 0412911-A 2 13-FEB-1991;

PLANT GENETIC SYSTEMS, N.V.

FEATURES  
source  
1..340  
Location/Qualifiers

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 106 a 73 c 92 g 69 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 340;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 42 TCAGAAGTATCAGCGACCTCCACC 65

## RESULT 5

BARARSTA BARARSTA 474 bp DNA BCT 23-JUN-1996  
LOCUS  
DEFINITION Bacillus amyloliquefaciens barstar gene.  
ACCESSION X15545  
VERSION X15545.1  
KEYWORDS GI:1155006  
SOURCE barstar; ribonuclease inhibitor.  
ORGANISM Bacillus amyloliquefaciens.

REFERENCE  
1 (bases 1 to 436)

AUTHORS Bacteria; Firmicutes; Bacillus/Clostridium group;

TITLE Bacillus/Staphylococcus group; Bacillus.

JOURNAL Hartley, R.W.

MEDLINE Barnase and barstar. Expression of its cloned inhibitor permits

REFERENCE expression of a cloned ribonuclease

89012012 J. Mol. Biol. 202 (4), 913-915 (1988)

2 (bases 1 to 474)

AUTHORS Hartley, R.

TITLE Direct Submission

JOURNAL Submitted (14-JAN-1996) R.Hartley, LCDB/NIDDK, NIH, Bethesda, 20892  
USA, email:hartley@helix.nih.gov  
REMARK Revised by author  
COMMENT On Jan 15, 1996 this sequence version replaced gi:39311.  
See also acc# X12871.

FEATURES  
source  
1..474  
Location/Qualifiers  
/organism="Bacillus amyloliquefaciens"  
/db\_xref="taxon:1390"  
/clones="pMT311"  
94..99  
/note="10 region"  
109..119  
/note="pot. ribosome binding site"  
124..396  
/note="barstar (AA 1 - 90)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA33551.1"  
/db\_xref="GI:39312"  
/db\_xref="SWISS-PROT:P11540"  
/translation="MKKAVINGEQIRISDLHQLKKELALPEYIGENLDALMDCLRG  
WVEYPLVLEWRQEQSKLTENGAEVQLVFREAKAGCDITILS"

BASE COUNT 154 a 104 c 123 g 93 t  
ORIGIN

Query Match 100.0%; Score 24; DB 1; Length 474;  
Best Local Similarity 100.0%; Pred. No. 0.098;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 155 TCAGAAGTATCAGCGACCTCCACC 178

RESULT 6  
A71435  
LOCUS A71435 4032 bp DNA circular PAT 07-MAY-1999  
DEFINITION Sequence 5 from Patent WO9810081.  
ACCESSION A71435  
VERSION A71435.1 GI:4775048  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified

REFERENCE 1 (bases 1 to 4032)  
AUTHORS Michiels,F. and Williams,M.  
TITLE IMPROVED BARSTAR GENE  
JOURNAL Patent: WO 9810081-A 5 12-MAR-1998;  
MICHELIS FRANK (BE)

FEATURES  
source  
1..4032  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 1072 a 968 c 963 g 1029 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4032;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 3435 TCAGAAGTATCAGCGACCTCCACC 3458

RESULT 7  
AR007527/c  
LOCUS AR007527 4808 bp DNA PAT 04-DEC-1998  
DEFINITION Sequence 17 from patent US 5750867.  
ACCESSION AR007527

VERSION AR007527.1 GI:3967011  
KEYWORDS  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4808)  
AUTHORS Williams,M. and Leemans,J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5750867-A 17 12-MAY-1998;  
FEATURES Location/Qualifiers  
source 1..4808  
/organism="unknown"

BASE COUNT 1370 a 1063 c 1038 g 1333 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4808;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 8  
AR084093/c  
LOCUS AR084093 4808 bp DNA PAT 01-SEP-2000  
DEFINITION Sequence 17 from patent US 5977433.  
ACCESSION AR084093  
VERSION AR084093.1 GI:10010864  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.

REFERENCE 1 (bases 1 to 4808)  
AUTHORS Williams,M. and Leemans,J.  
TITLE Maintenance of male-sterile plants  
JOURNAL Patent: US 5977433-A 17 02-NOV-1999;  
FEATURES Location/Qualifiers  
source 1..4808  
/organism="unknown"

BASE COUNT 1370 a 1063 c 1038 g 1333 t  
ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4808;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||  
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 9  
AX172441/c  
LOCUS AX172441 4832 bp DNA PAT 03-JUL-2001  
DEFINITION Sequence 2 from Patent WO0141558.  
ACCESSION AX172441  
VERSION AX172441.1 GI:14597553  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequence.

REFERENCE 1 (bases 1 to 4832)  
AUTHORS de Both,G. and de Beuckeleer,M.  
TITLE Hybrid winter oilseed rape and methods for producing same  
JOURNAL Patent: WO 0141558-A 2 14-JUN-2001;  
Aventis CropScience N.V. (BE)

FEATURES  
source  
1..4832  
/organism="synthetic construct"

misc\_feature /db\_xref="taxon:32630"  
 1883. .4065 /note="T-DNA of plasmid pTHW118"  
 BASE COUNT 1528 a 883 c 932 g 1488 t 1 others  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 4832;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 3222 TCAGAAGTATCAGCGACCTCCACC 3199

RESULT 10  
 AX127748  
 LOCUS AX127748 5865 bp DNA PAT 15-MAY-2001  
 DEFINITION Sequence 1 from Patent WO0131042.  
 ACCESSION AX127748  
 VERSION AX127748.1 GI:14134395

KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM  
 REFERENCE 1 (bases 1 to 5865)  
 AUTHORS Weston,B. and de Beuckeleer,M.  
 TITLE Male-sterile brassica plants and methods for producing same  
 JOURNAL Patent: WO 0131042-A 1 03-MAY-2001;  
 FEATURES Aventis CropScience N.V. (BE)  
 source Location/Qualifiers  
 1. 5865  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="T-DNA of plasmid pCol13"

BASE COUNT 1849 a 1095 c 1149 g 1772 t  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 5865;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 5249 TCAGAAGTATCAGCGACCTCCACC 5272

RESULT 11  
 AR007512/c  
 LOCUS AR007512 6555 bp DNA PAT 04-DEC-1998  
 DEFINITION Sequence 2 from patent US 5750867.  
 ACCESSION AR007512  
 VERSION AR007512.1 GI:3966996

KEYWORDS  
 SOURCE Unknown.  
 ORGANISM  
 REFERENCE 1 (bases 1 to 6555)  
 AUTHORS Williams,M. and Leemans,J.  
 TITLE Maintenance of male-sterile plants  
 JOURNAL Patent: US 5750867-A 2 12-MAY-1998;  
 FEATURES Location/Qualifiers  
 source 1. .6555  
 /organism="unknown"

BASE COUNT 1690 a 1611 c 1584 g 1670 t  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 6555;  
 Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 993 TCAGAAGTATCAGCGACCTCCACC 970

RESULT 12  
 AR084078/c  
 LOCUS AR084078 6555 bp DNA PAT 01-SEP-2000  
 DEFINITION Sequence 2 from patent US 5977433.  
 ACCESSION AR084078  
 VERSION AR084078.1 GI:10010849  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM

REFERENCE 1 (bases 1 to 6555)  
 AUTHORS Williams,M. and Leemans,J.  
 TITLE Maintenance of male-sterile plants  
 JOURNAL Patent: US 5977433-A 2 02-NOV-1999;  
 FEATURES Location/Qualifiers  
 source 1. .6555  
 /organism="unknown"

BASE COUNT 1690 a 1611 c 1584 g 1670 t  
 ORIGIN

Query Match 100.0%; Score 24; DB 6; Length 6555;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||  
 Db 993 TCAGAAGTATCAGCGACCTCCACC 970

RESULT 13  
 A71433  
 LOCUS A71433 273 bp DNA PAT 07-MAY-1999  
 DEFINITION Sequence 3 from Patent WO9810081.  
 ACCESSION A71433  
 VERSION A71433.1 GI:4775046  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unclassified.

REFERENCE 1 (bases 1 to 273)  
 AUTHORS Michiels,F. and Williams,M.  
 TITLE IMPROVED BARSTAR GENE  
 JOURNAL Patent: WO 9810081-A 3 12-MAR-1998;  
 FEATURES MICHIELS FRANK (BE)  
 source Location/Qualifiers  
 1. .273  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 1. .>273  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAB42578.1"  
 /db\_xref="GI:4775047"

BASE COUNT 62 a 84 c 85 g 42 t  
 ORIGIN  
 GWVEYPLVLEWRQFQSKLTENGAEVLFQVFEAKAGECDDITILS"  
 /translation="MAKKAVINGEQIRISIDLHQTLKKELALPEVYGENLDALWDCLT"

Query Match 86.7%; Score 20.8; DB 6; Length 273;  
 Best Local Similarity 91.7%; Pred. No. 4.6;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 |||||



```
Db 35 TCAGGAGCATCAGCGACCTCCACC 58

RESULT 14
A71436
LOCUS A71436 563 bp DNA
DEFINITION Sequence 6 from Patent WO9810081.
ACCESSION A71436
VERSION A71436.1 GI:4775049
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 563)
AUTHORS Michiels,F. and Williams,M.
TITLE IMPROVED BARSTAR GENE
JOURNAL Patent: WO 9810081-A 6 12-MAR-1998;
MICHIELS FRANK (BE)
FEATURES
    source
        location/Qualifiers
            1..563
                /organism="unidentified"
                /db_xref="taxon:32644"
BASE COUNT 152 a 133 c 145 g 133 t
ORIGIN

Query Match 86.7%; Score 20.8; DB 6; Length 563;
Best Local Similarity 91.7%; Pred. No. 4.8;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
||||| ||||| ||||| ||||| |||||
Db 37 TCAGGAGCATCAGCGACCTCCACC 60

RESULT 15
A71437
LOCUS A71437 5349 bp DNA
DEFINITION Sequence 7 from Patent WO9810081.
ACCESSION A71437
VERSION A71437.1 GI:4775050
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 5349)
AUTHORS Michiels,F. and Williams,M.
TITLE IMPROVED BARSTAR GENE
JOURNAL Patent: WO 9810081-A 7 12-MAR-1998;
MICHIELS FRANK (BE)
FEATURES
    source
        location/Qualifiers
            1..5349
                /organism="unidentified"
                /db_xref="taxon:32644"
BASE COUNT 1339 a 1233 c 1290 g 1487 t
ORIGIN

Query Match 86.7%; Score 20.8; DB 6; Length 5349;
Best Local Similarity 91.7%; Pred. No. 5.3;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
||||| ||||| ||||| ||||| |||||
Db 4004 TCAGGAGCATCAGCGACCTCCACC 4027

Search completed: February 25, 2002, 18:00:23
Job time: 18431 sec
```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 25, 2002, 18:17:34 ; Search time 716.55 seconds  
(without alignments)  
28,715 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24

Sequence: 1 tcagaagatcagcgacctccacc 24

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	22	PCR primer MDB251
2	24	100.0	270	19	Wild type barstar
3	24	100.0	323	19	Barstar coding seq
4	24	100.0	340	12	Clai-HindII fragme
5	24	100.0	1303	17	Plasmid pTS88 (Eco
c 6	24	100.0	2275	22	Oligonucleotide #1
c 7	24	100.0	3544	17	PTS200 contg. p35S
c 8	24	100.0	3544	20	Nucleotide sequenc
c 9	24	100.0	4032	19	Plasmid pmv71. Sy
c 10	24	100.0	4808	15	Restriction fragme
c 11	24	100.0	4832	22	Nucleotide sequenc

12	24	100.0	4896	17	AAT08976
13	24	100.0	4896	20	AA15631
14	24	100.0	5864	17	AAT39339
15	24	100.0	5865	22	AAD06990
c 16	24	100.0	6555	15	AAQ53874
c 17	24	100.0	7492	22	AAF86441
18	20.8	86.7	563	19	AAV23235
19	20.8	86.7	563	19	AAV23235
20	20.8	86.7	5349	19	AAV23239
21	20.8	86.7	5611	19	AAV23242
c 22	17.6	73.3	245	22	AAI52182
c 23	17.6	73.3	469	22	AAI39127
c 24	17.6	73.3	3534	20	AAZ41289
c 25	17.6	73.3	4721	22	AAI59601
c 26	17.6	73.3	4753	22	AAI57815
c 27	17.2	71.7	384	22	AAI20952
c 28	17.2	71.7	384	22	AAI46196
c 29	17.2	71.7	384	22	AAI06663
c 30	17.2	71.7	450	22	AAI11739
c 31	17.2	71.7	450	22	AAI33046
c 32	17.2	71.7	450	22	AAI01667
c 33	17.2	71.7	543	21	AAC34918
c 34	17.2	71.7	616	21	AAC50006
c 35	17.2	71.7	10223	19	AAV52206
c 36	17.2	71.7	160552	22	AAD02697
c 37	17	70.8	1842	22	AAC83227
c 38	16.8	70.0	3831	19	AAV52424
c 39	16.6	69.2	154	21	AAA89494
c 40	16.6	69.2	609	22	AAH67560
c 41	16.6	69.2	1239	22	AAH45076
c 42	16.6	69.2	1292	22	AAD08025
c 43	16.6	69.2	1371	22	AAD08026
c 44	16.6	69.2	1650	22	AAF33121
c 45	16.6	69.2	1866	17	AAT41492

ALIGNMENTS

RESULT 1

AD06998	ID	AAD06998 standard; DNA; 24 BP.
XX	AC	AAD06998;
XX	DT	06-AUG-2001 (first entry)
XX	DE	PCR primer MDB251 to generate the flanking region of elite event MS-B2.
XX	MS-B2	elite event; transgenic Brassica plant; transformation event;
KW	male-sterility	gene; PCR primer; thermal asymmetric interlaced;
KW	TAIL; ss.	
XX	OS	Bacillus amyloliquefaciens.
XX	PN	WO200131042-A2.
XX	PD	03-MAY-2001.
XX	PF	26-OCT-2000; 2000WO-EF10680.
XX	PR	29-OCT-1999; 99US-0430497.
XX	(AVET)	AVENTIS CROPS SCIENCE NV.
XX	Weston B,	De Beuckeleer M;
XX	WPI;	2001-300517/31.

Transgenic Brassica plants, seeds, cells or tissues, characterized by harboring specific transformation events, particularly by presence of male-sterility gene, at specific location in its genome -

PS Example 3; Page 28; 53pp; English.

CC The present invention relates to a transgenic Brassica plant or its

CC seed, cells or tissues, characterised by harbouring a specific

CC transformation event, particularly by the presence of a male-sterility

CC gene, at a specific location in the Brassica genome. Transgenic

CC Brassica plant is useful for producing a hybrid seed by crossing the

CC transgenic plant with a male-fertile Brassica plant and harvesting the

CC hybrid seed from the transgenic Brassica plant.

CC The present sequence is primary thermal interlaced (TAIL)-PCR primer

CC M8B8 used to left (3') border flanking region of elite event MS-B2.

CC This primer corresponds to position 5249-5272 of plasmid pTC0113.

XX

SQ Sequence 24 BP; 7 A; 9 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 24; DB 22; Length 24;

Best Local Similarity 100.0%; Pred. No. 0.036;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24

DB 1 tcagaagtatcagcgacctccacc 24

|||||

RESULT 2

AAV23236

ID AAV23236 standard; DNA; 270 BP.

XX

AC AAV23236;

XX

DT 17-JUL-1998 (first entry)

XX

DE Wild type barstar DNA.

XX

KW Barstar; barnase inhibitor; fertility restoration;

KX male-sterile line; ds.

XX

OS Bacillus amyloliquefaciens.

XX

PH Key Location/Qualifiers

FT CDS 1..273

FT /\*tag= a

FT /product= barstar

FT /note= "stop codon not given"

XX

PN W09810081-A2.

XX

PD 12-MAR-1998.

XX

XX 01-SEP-1997; 97WO-EP04739.

XX

PR 03-SEP-1996; 96EP-0202446.

XX

PA (PLBZ ) PLANT GENETIC SYSTEMS NV.

XX

PI Michiels F, Williams M;

XX

DR WPI; 1998-193630/17.

DR

DR P-PSDB; AAW53344.

XX

PT DNA encoding an improved barstar protein - used to restore fertility

PT In male-sterile plant lines

XX

PS Claim 8; Pages 34-35; 54pp; English.

XX

CC The present sequence was used in the preparation of an improved

CC Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which

CC can be used to restore fertility to male-sterile lines.

CC The DNA sequence encoding the improved barstar, leads to increased

CC barstar production in tapetum cells, due to improved translation,

CC and possibly protein stability.

XX

SQ Sequence 270 BP; 80 A; 58 C; 74 G; 58 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 270;

Best Local Similarity 100.0%; Pred. No. 0.048;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcgacctccacc 24

DB 32 tcagaagtatcagcgacctccacc 55

|||||

RESULT 3

AAV60977

ID AAV60977 standard; DNA; 323 BP.

XX

AC AAV60977;

XX

DT 03-DEC-1998 (first entry)

XX

DE Barstar coding sequence.

XX

KW Barnase; barstar; IPCR; inverse polymerase chain reaction; phenotype;

KX transgenic plant; hybrid seed; male sterile plant; active enzyme;

XX

OS Synthetic.

XX

PH Key Location/Qualifiers

FT CDS 1..273

FT /\*tag= a

FT /product= "barstar"

XX

PN W09837211-A1.

XX

PD 27-AUG-1998.

XX

PF 20-FEB-1998; 98WO-GB00542.

XX

PR 21-FEB-1997; 97GB-0003681.

XX

PA (GENE-) GENE SHEARS PTY LTD.

XX

PI Betzner AS, Huttner E, Paul W, Perez P;

XX

DR WPI; 1998-467572/40.

DR

DR P-PSDB; AAW71704.

XX

PT Production of transgenic plants having a desired phenotype - by

PT using a pair of parent plants which each produce a polypeptide which

PT complement each other when crossed

XX

PS Example 1; Fig 1C; 58pp; English.

XX

CC The present invention describes a pair of parent plants for producing

CC seeds comprising: (a) a first parent plant containing at least 1 gene

CC sequence encoding a polypeptide or protein A, and (b) a second parent

CC plant containing at least 1 gene sequence encoding a polypeptide or

CC protein B; where the polypeptides A and B, when expressed in separate

CC plants, do not form an active enzyme, a regulatory protein or protein

CC which affects the functionality and/or viability and/or the structural

CC integrity of a cell, but when expressed in the same plant do form an

CC active enzyme, regulatory protein, or protein which affects the

CC structural integrity of a cell. Also described is a method for producing

CC a plant having a desired phenotype by virtue of an active enzyme, a

CC regulatory protein or a protein which affects the structural integrity

CC of a cell comprising crossing a first line with a second line where the

CC first line contains one or more gene sequences encoding a polypeptide or

CC protein A but which line does not have the desired phenotype and where

CC the second line contains one or more gene sequences encoding a

CC polypeptide or protein B which is complementary to the polypeptide or

CC protein A but which line does not have the desired phenotype. The method

CC can be used for producing plants having altered phenotypes, e.g. male-

CC sterility, embryoless seeds, altered biochemical (e.g. fatty acid)  
 CC composition or herbicide resistance. The present sequence encodes  
 CC barstar which is used in an example from the present invention.  
 XX  
 SQ Sequence 323 BP; 98 A; 69 C; 87 G; 69 T; 0 other;

Query Match 100.0%; Score 24; DB 19; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 0.049;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagaagtatcagcgcctccacc 24  
 |||||  
 Db 32 tcagaagtatcagcgcctccacc 55

RESULT 4  
 ID AAQ10460 standard; DNA; 340 BP.  
 XX  
 AC AAQ10460;  
 DT 16-APR-1991 (first entry)  
 XX  
 DE ClaI-HindII fragment contg barstar gene.  
 XX  
 KW Inhibitor; extracellular ribonuclease; barnase; fertility; ss.  
 XX  
 OS Bacillus amyloliquefaciens.  
 PN EP412911-A.  
 XX  
 PD 13-FEB-1991.  
 XX  
 PF 09-AUG-1990; 90EP-0402281.  
 XX  
 PR 10-AUG-1989; 89EP-0402270.  
 XX  
 PA (PLAN-) PLANT GENETIC SYST.  
 XX  
 PI Mariani C, Leemans J, De Greef W;  
 XX  
 DR WPT; 1991-046026/07.  
 XX

Cell of fertility restored plant - in which nuclear genome in  
 transformed with foreign DNA sequence neutralising activity of  
 another prod.  
 XX  
 PS Disclosure; Fig 2; 25pp; English.  
 XX  
 CC The barstar gene encodes and inhibitor of barnase, which degrades  
 CC RNA molecules by hybridising the bond after a guanine residue.  
 CC The gene is used, in a chimaeric sequence, to restore fertility in  
 CC plants transformed with the gene. The promoter cassette PTA29  
 CC (EP-401194) is fused in frame with the initiating ATG.  
 XX  
 SQ Sequence 340 BP; 106 A; 73 C; 92 G; 69 T; 0 other;

Query Match 100.0%; Score 24; DB 12; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 0.05;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcagaagtatcagcgcctccacc 24  
 |||||  
 Db 42 tcagaagtatcagcgcctccacc 65

RESULT 5  
 AAT39337  
 ID AAT39337 standard; DNA; 1303 BP.  
 XX  
 AC AAT39337;

XX 22-JAN-1997 (first entry)  
 DT  
 XX Plasmid pTS88 (EcoRI-HindIII fragment).  
 DE  
 XX Plasmid pTS88; male sterile; barstar; barnase; ribonuclease;  
 KW transgenic plant; rice; Oryza sativa; maize; corn; Zea mays; ds.  
 XX  
 OS Synthetic.  
 XX  
 PH Key Location/Qualifiers  
 FT misc\_feature 1..35  
 FT /tag= a  
 FT /label= pGEM2  
 FT /note= "polylinker of pGEM2"  
 FT 36..694  
 FT /tag= b  
 FT /label= P35S  
 FT /function= 35S promoter of cauliflower mosaic virus  
 FT strain CM1841  
 FT CDS 695..967  
 FT /tag= c  
 FT /label= barstar  
 FT /product= Bacillus amyloliquefaciens barstar  
 FT 968..1287  
 FT /tag= d  
 FT /label= 3'g7  
 FT /function= region containing polyadenylation signal  
 FT Of gene 7 og Agrobacterium T-DNA  
 FT misc\_feature 1288..1303  
 FT /tag= e  
 FT /label= pGEM2  
 FT /note= "polylinker of pGEM2"  
 FT  
 XX WO9626283-A1.  
 XX  
 PD 29-AUG-1996.  
 XX  
 PF 21-FEB-1996; 96WO-EP00722.  
 XX  
 PR 21-FEB-1995; 95EP-0400364.  
 XX  
 PA (PLB2 ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Botterman J, Cornelissen M, Michiels F;  
 XX  
 DR WPI; 1996-402373/40.  
 XX  
 PT Prodn. of male sterile plants by transforming with a chimaeric  
 PT construct - comprising a male sterility DNA e.g. barnase and a  
 PT co-regulating gene, e.g. barstar, into the nuclear genome, useful  
 PT for generating hybrid cultivars  
 XX  
 PS Example 1; Page 38; 56pp; English.  
 XX  
 CC The HindIII-EcoRI fragment (AAT39337) of plasmid pTS88 contains  
 CC barstar DNA under control of a 35S promoter. The plasmid was  
 CC used with pTS174 (see also AAT39336) contg. barnase DNA under  
 CC control of the stamen-specific promoter Ei to produce male sterile  
 CC rice cv. Kochihibiki transgenic plants, and with plasmid pVEI36  
 CC (see also AAT39338) contg. barnase DNA under control of the stamen-  
 CC specific PCA55 promoter to produce male sterile maize plants.  
 CC Expression of barnase (a ribonuclease) in the stamen leads to male  
 CC sterility. Constitutive expression of barstar counteracts possible  
 CC low level expression of barnase DNA in non-stamen tissue.  
 XX  
 SQ Sequence 1303 BP; 415 A; 287 C; 275 G; 326 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 1303;  
 Best Local Similarity 100.0%; Pred. No. 0.059;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 Db 726 tcagaagtatcagcgacctccacc 749

## RESULT 6

AAF86440/c  
 ID AAF86440 standard; DNA; 2275 BP.  
 XX  
 AC AAF86440;  
 XX  
 DT 25-JUN-2001 (first entry)  
 XX  
 DE Oligonucleotide #1: SEQ ID 4.  
 XX  
 KW Male sterile plant; RNAase inhibitor; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200124616-A1.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 12-SEP-2000; 2000WO-JP06222.  
 XX  
 PR 30-SEP-1999; 99JP-0279307.  
 XX  
 PA (NTSB ) JAPAN TOBACCO INC.  
 XX  
 PI Hamada K, Nakakido F;  
 XX  
 DR WPI; 2001-266212/27.  
 XX

PT Method for producing male sterile rice and maize by inserting RNAse  
 gene and RNAse inhibitor genes with promoters into the plant genome -  
 XX  
 PS Disclosure; Page 17-19; 29pp; Japanese.

CC The present invention relates to a method for producing male sterile  
 CC plants. The method comprises inserting a promoter fragment upstream of an  
 CC RNAse gene and a second promoter, upstream of an RNAse inhibitor protein  
 CC gene and inserting it into the plant genome. The method is useful for  
 CC producing male sterile tobacco, lettuce and rapeseed plants, but  
 CC preferably rice and maize. The present sequence is an oligonucleotide  
 CC used in the method of the present invention.

XX Sequence 2275 BP; 604 A; 496 C; 496 G; 679 T; 0 other;

Query Match 100.0%; Score 24; DB 22; Length 2275;  
 Best Local Similarity 100.0%; Pred. No. 0.063;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 Db 533 TCAGAAGTATCAGCGACCTCCACC 510

## RESULT 7

AAT17246/c  
 ID AAT17246 standard; DNA; 3544 BP.  
 XX  
 AC AAT17246;  
 XX

DT 12-AUG-1996 (first entry)

XX PTS200 contg. P35S-bar-3'nos and PCA55-barstar-3'nos.  
 XX

XX Cl; Cl-S; transgenic plant; male sterility; colour-linked restorer;  
 KW anthocyanin; aleurone; B-peru; bar; barstar; bar\*; ds.  
 XX Synthetic.  
 OS  
 XX

FT FH Key Location/Qualifiers  
 FT FT misc\_signal  
 FT 3227..3504  
 FT /\*tag= a  
 FT /label= 3'nos  
 FT /note= "3' regulatory sequence contg. the  
 FT polyadenylation signal of the nopaline  
 FT synthase gene of Agrobacterium T-DNA"  
 FT CDS  
 FT 2675..3226  
 FT /\*tag= b  
 FT /label= bar  
 FT /note= "coding region of bar gene of  
 FT Streptomyces hygroscopicus"  
 FT promoter  
 FT 1841..2674  
 FT /\*tag= c  
 FT /label= P35S  
 FT /note= "35S promoter of Cauliflower Mosaic Virus"  
 FT complement (626..1803)  
 FT /\*tag= d  
 FT /label= PCA55  
 FT /note= "promoter of CA55 gene of Zea mays"  
 FT complement (353..625)  
 FT /\*tag= e  
 FT /label= barstar  
 FT /note= "coding region of barstar gene of  
 FT Bacillus amyloliquefaciens"  
 FT complement (30..352)  
 FT /\*tag= f  
 FT /label= 3'nos  
 FT /note= "3' regulatory sequence contg. the  
 FT polyadenylation signal of the nopaline  
 FT synthase gene of Agrobacterium T-DNA"  
 FT XX  
 PN WO534634-A2.  
 XX  
 PD 21-DEC-1995.  
 XX  
 PF 06-JUN-1995; 95WO-EP02157.  
 XX  
 PR 06-JUN-1994; 94US-0254776.  
 XX  
 PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Kriebbers E, Leemans J, Williams M;  
 XX  
 DR WPI; 1996-049664/05.  
 XX  
 PT Transgenic plants contg. male sterility and colour-linked restorer  
 PT genotypes - used for prodn. of male sterile seeds identifiable from  
 PT their colour, also new truncated anthocyanin regulatory genes and  
 PT aleurone specific promoters  
 XX  
 PS Disclosure; Page 72-74; 104pp; English.  
 XX  
 CC Shortened forms of anthocyanin (An) regulatory genes (ARG) are easier  
 CC to manipulate than the complete gene and still provide An prodn.  
 CC The Cl gene (and the Cl-S gene) can be considerably shortened  
 CC while still retaining, under appropriate conditions, its  
 CC capability of conditioning anthocyanin prodn. in the aleurone  
 CC of seeds of cereal plants such as corn. A pref. shortened Cl gene  
 CC is comprised in pCOL9 (see AAT08975). The full Cl gene sequence is  
 CC given in AAT08973.  
 CC A truncated B-peru gene (presumed sequence = AAT08674; actual  
 CC sequence = AAT08977) may also be used.  
 CC Plasmids PTS256 (AAT08976) and PTS200 (AAT17246) were used in the  
 CC construction of vectors comprising the Cl and B-peru genes  
 CC as well as male-sterility gene and a selectable marker gene.  
 XX  
 SQ Sequence 3544 BP; 1011 A; 847 C; 767 G; 919 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 3544;  
 Best Local Similarity 100.0%; Pred. No. 0.067;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
 ||||||||||||||||||  
 Db 594 TCAGAAGTATCAGCGACCTCCACC 571

RESULT 8  
 AAX15632/c  
 ID AAX15632 standard; DNA; 3544 BP.  
 XX AC AAX15632;  
 XX AC AAX15632;

DT 07-MAY-1999 (first entry)

DE Nucleotide sequence of an EcoRI-HindIII fragment of plasmid pTS256.

KW C1 gene; maize; male-sterile corn line; anthocyanin production;  
 KW pTS256; ds.  
 XX Synthetic.

OS US5880331-A.

PN 09-MAR-1999.

PD 07-JUN-1995; 95US-0485139.

PF 07-JUN-1995; 95US-0485139.

PR 06-JUN-1994; 94US-0254776.

XX (PUBZ ) PLANT GENETIC SYSTEMS NV.

PA Krebbers E, Leemans J, Williams M;

PI WPI; 1999-204053/17.

DR Process for maintaining a male-sterile corn line - using male

PT sterile plants lacking functional regulatory gene for anthocyanin

PT production

PS Example 2; Columns 37-40; 35pp; English.

XX The present sequence represents the nucleotide sequence of an

CC EcoRI-hindIII fragment of plasmid pTS256, comprising the chimeric

CC gene PCA55-barstar-3'nos. It is used in the course of the invention. The

CC specification describes a process for maintaining a male-sterile corn

CC line, using male sterile parent plants lacking a functional gene for

CC anthocyanin production, and a maintainer corn line comprising male

CC fertile parent plants containing foreign DNA comprising a restorer gene

CC and an active regulatory protein gene. By using the anthocyanin gene,

CC the colour of the male-sterile plants will differ from that of the

CC male fertile plants. This will enable the seeds harvested from the

CC plants to be easily separated into those that will grow into

CC male-fertile plants and those that will grow into male-sterile plants.

XX Sequence 3544 BP; 1011 A; 847 C; 767 G; 919 T; 0 other;

SQ Query Match 100.0%; Score 24; DB 20; Length 3544;

Best Local Similarity 100.0%; Pred. No. 0.067;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 17-JUL-1998 (first entry)

XX Plasmid pMW71.

XX Barstar; barnase inhibitor; fertility restoration;

KW male-sterile line; plasmid pMW71; circular; ds.

XX Synthetic.

XX Key Location/Qualifiers

FT promoter

FT /tag= a

FT /note= "promoter region of rice actin gene -

FT contains an intron in the leader"

FT CDS 3401..3676

FT /tag= b

FT /product= barstar

FT 3'UTR 3677..4003

FT /tag= c

FT /note= "region containing 3' untranslated end of

FT the nopaline synthase gene of Agrobacterium

FT T-DNA"

XX WO9810081-A2.

XX 12-MAR-1998.

XX 01-SEP-1997; 97WO-EP04739.

XX 03-SEP-1996; 96EP-0202446.

XX (PLBZ ) PLANT GENETIC SYSTEMS NV.

PA Michiels F, Williams M;

PI WPI; 1998-193630/17.

DR DNA encoding an improved barstar protein - used to restore fertility

PT in male-sterile plant lines

XX Example 3; Pages 37-39; 54pp; English.

XX The present sequence was used in the preparation of an improved

CC Bacillus amyloliquefaciens barstar, i.e. barnase inhibitor, which

CC can be used to restore fertility to male-sterile lines.

CC The DNA sequence encoding the improved barstar, leads to increased

CC barstar production in tapetum cells, due to improved translation,

CC and possibly protein stability.

XX Sequence 4032 BP; 1072 A; 968 C; 963 G; 1029 T; 0 other;

SQ Query Match 100.0%; Score 24; DB 19; Length 4032;

Best Local Similarity 100.0%; Pred. No. 0.068;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24

||||||||||||||||||||

Db 3435 tcagaagtatcagcgacctccacc 3458

RESULT 10

AAQ53889/c

ID AAQ53889 standard; DNA; 4808 BP.

XX AAQ53889;

XX 27-JUN-1994 (first entry)

XX Restriction fragment of construct carrying plant maintainer gene.

DE Maintainer gene; sterile; sterility; homogenous population; hybrid;

XX seed; fertility restorer gene; pollen lethality gene; ss.

KW

```

XX OS Synthetic.
XX FH Key Location/Qualifiers
FT FT misc_signal complement (18..401)
FT FT /*tag= a
FT FT /label= 3' nos.
FT FT /note= "3' regulatory sequence containing the
FT FT polyadenylation site derived from
FT FT Agrobacterium T-DNA nopaline synthase gene."
FT FT CDS complement (402..737)
FT FT /*tag= b
FT FT /label= Barnase.
FT FT /note= "Coding region of the barnase gene of
FT FT Bacillus amyloliquefaciens."
FT FT promoter complement (738..1944)
FT FT /*tag= c
FT FT /label= P2M13.
FT FT /note= "Promoter region of the Zm 13 gene of Zea
FT FT mays."
FT FT misc_signal complement (1945..2281)
FT FT /*tag= d
FT FT /label= 3' nos.
FT FT CDS complement (2282..2554)
FT FT /*tag= e
FT FT /label= Barstar.
FT FT /note= "Coding region of the barstar gene of
FT FT Bacillus amyloliquefaciens."
FT FT promoter complement (2555..3099)
FT FT /*tag= f
FT FT /label= PTA29.
FT FT /note= "Promoter region of the TA29 gene of
FT FT Nicotiana tabacum."
FT FT promoter 3100..3932
FT FT /*tag= g
FT FT /label= 3S3.
FT FT /note= "3S3" promoter sequence derived from
FT FT cauliflower mosaic virus isolate CabBB-J1."
FT FT CDS 3933..4484
FT FT /*tag= h
FT FT /note= "Coding region of the phosphinothricin
FT FT acetyltransferase gene."
FT FT misc_signal 4485..4763
FT FT /*tag= i
FT FT /label= 3' nos.
FT FT WO9325695-A.
FT FT 23-DEC-1993.
FT FT 11-JUN-1993; 93WO-EP01489.
FT FT 12-JUN-1992; 92US-0899072.
FT FT 03-NOV-1992; 92US-0970840.
FT FT (PLBZ ) PLANT GENETIC SYSTEMS NV.
FT FT Leemans J, Williams M;
FT FT WPI; 1994-007552/01.
FT FT Maintainer gene for maintenance of male-sterile plants -
FT FT comprises fertility-restorer gene and pollen-lethality gene
FT FT Example 2; Page 73-75; 87pp; English.
FT FT A maintainer gene of plants, pref. a foreign chimeric gene,
FT FT comprises (a) a fertility restorer gene which comprises a fertility
FT FT restorer DNA and (ii) a restorer promoter capable of
FT FT directing the expression of the fertility restorer DNA and (b) a
FT FT pollen lethality gene that is selectively expressed in microspores
FT FT and/or pollen of the plant to prevent the production of functional
FT FT pollen and which comprises (i) a pollen lethality DNA and (ii) a

```

```

CC pollen specific promoter capable of directing expression of the
CC pollen lethality DNA. Plants transformed with this DNA (maintainer
CC plants) can be used to maintain a homogenous population of male
CC sterile plants for the production of hybrid seed. This sequence is
CC an EcoRI-HindIII restriction fragment of the construct designated
CC pTS218 and comprises the maintainer gene described.
XX SQ Sequence 4808 BP; 1370 A; 1064 C; 1037 G; 1333 T; 4 other;

Query Match 100.0%; Score 24; DB 15; Length 4808;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
|||||
Db 2529 TCAGAAAGTATCAGCGACCTCCACC 2506

RESULT 11
AAH25423/c
ID AAH25423 standard; DNA; 4832 BP.
XX AC AAH25423;
XX DT 22-AUG-2001 (first entry)
XX DE Nucleotide sequence of plasmid pTHW118.
XX KW Transgenic plant; winter oilseed rape; hybrid seed; male-sterility gene;
XX KW fertility restorer gene; barstar gene; ss.
OS Synthetic.
OS Streptomyces hygroscopicus.
OS Arabidopsis thaliana.
OS Bacillus amyloliquefaciens.
OS Nicotiana tabacum.
XX FH Key Location/Qualifiers
FT FT misc_feature 1..25
FT FT /*tag= a
FT FT /note= "right border repeat from TL-DNA from pTIB6S3"
FT FT misc_feature 26..53
FT FT /*tag= b
FT FT /note= "synthetic polylinker derived sequences"
FT FT misc_feature 54..90
FT FT /*tag= c
FT FT /note= "residual sequence from TL-DNA at right
FT FT border repeat"
FT FT misc_feature 91..97
FT FT /*tag= d
FT FT /note= "synthetic polylinker derived sequences"
FT FT 3'UTR complement (98..309)
FT FT /*tag= e
FT FT /note= "3' UTR from TL-DNA gene 7 of pTIB6S3"
FT FT misc_feature 310..330
FT FT /*tag= f
FT FT /note= "synthetic polylinker derived sequences"
FT FT CDS complement (331..882)
FT FT /*tag= g
FT FT /note= "Streptomyces hygroscopicus bialaphos
FT FT resistance (bar) gene"
FT FT promoter complement (883..2608)
FT FT /*tag= h
FT FT /note= "atSIA ribulose-1,5-biphosphate carboxylase
FT FT small subunit gene from Arabidopsis thaliana"
FT FT misc_feature 2609..2658
FT FT /*tag= i
FT FT /note= "synthetic polylinker derived sequences"
FT FT 3'UTR complement (2659..2919)
FT FT /*tag= j
FT FT /note= "taql fragment from 3' UTR of nopaline
FT FT synthase gene from T-DNA of pTIF37 and

```



```

FT  misc_feature      2920..2940      containing plant polyadenylation signals"
FT  /*tag= k
FT  /note= "synthetic polylinker derived sequences"
FT  2941..2980
FT  /*tag= l
FT  /note= "downstream of Bacillus amyloliquefaciens
FT  barstar coding region"
FT  CDS
FT  complement (2981..3253)
FT  /*tag= m
FT  /note= "Barstar gene coding region from Bacillus
FT  amyloliquefaciens"
FT  complement (3254..4762)
FT  /*tag= n
FT  /note= "anther-specific gene TA29 promoter from
FT  Nicotiana tabacum"
FT  4763..4807
FT  /*tag= o
FT  /note= "synthetic polylinker derived sequences"
FT  4808..4832
FT  /*tag= p
FT  /note= "left border repeat from TL-DNA from pTIB6S3"
XX  WO200141558-A1.
XX
XX  14-JUN-2001.
XX
XX  06-DEC-2000; 2000WO-EP12872.
XX
XX  08-DEC-1999; 99US-0457037.
XX
XX  (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX  De Both G, De Beuckeleer M;
XX
XX  WPI; 2001-381419/40.
XX
XX  Transgenic winter oilseed rape plants suited for producing hybrid seed
XX  with improved qualities, comprises a male-sterility gene and fertility
XX  restorer gene, integrated into the genome -
XX
XX  Example 1; Page 80-82; 98pp; English.
XX
XX  The specification describes a pair of transgenic winter oilseed rape
XX  plants suited for producing hybrid seed. One of the plants has an
XX  expression cassette comprising a male-sterility gene, and the other
XX  plant has an expression cassette comprising a fertility restorer gene,
XX  integrated into the genome. The fertility restorer gene is capable of
XX  preventing the activity of the male-sterility gene. The plant pair is
XX  useful for producing hybrid seed. Plants developed from the hybrid
XX  seed have agronomic performance, genetic stability and adaptability to
XX  different genetic backgrounds. The present sequence represents
XX  plasmid pTHW118. This plasmid comprises the barstar gene, which acts as
XX  a fertility restorer gene. The plasmid is used to create transgenic
XX  plants of the invention.
XX
XX  Sequence 4832 BP; 1528 A; 883 C; 932 G; 1488 T; 1 other;

Query Match      100.0%; Score 24; DB 22; Length 4832;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy  1 tcagaagtatcagcgacctccacc 24
    |||||||
Db  3222 TCAGAAGTATCAGCGACCTCCACC 3199

RESULT 12
AAT08976
ID  AAT08976 standard; DNA; 4896 BP.
XX
AC  AAT08976;

```

```

XX  12-AUG-1996 (first entry)
DT
XX  PTS256 contg. P35s-bar-3'nos and PTA29-barstar-3'nos.
DE
XX  C1; C1-S; transgenic plant; male sterility; colour-linked restorer;
KW  anthocyanin; aleurone; B-peru; bar; barstar; bar*; ds.
XX  Synthetic.
OS
XX  Key
FH  misc_signal
FT  complement (39..317)
FT  /*tag= a
FT  /label= 3'nos
FT  /note= "3' regulatory sequence contg. the
FT  polyadenylation signal of the nopaline
FT  synthase gene of Agrobacterium T-DNA"
FT  CDS
FT  complement (318..869)
FT  /*tag= b
FT  /label= bar
FT  /note= "coding region of bar gene of
FT  Streptomyces hygroscopicus"
FT  complement (870..1702)
FT  /*tag= c
FT  /label= P35S
FT  /note= "35S promoter of Cauliflower Mosaic Virus"
FT  complement (1740..2284)
FT  /*tag= d
FT  /label= PTA29
FT  /note= "promoter of TA29 gene of Nicotiana
FT  tabacum"
FT  CDS
FT  2285..2557
FT  /*tag= e
FT  /label= barstar
FT  /note= "coding region of barstar gene of
FT  Bacillus amyloliquefaciens"
FT  misc_signal
FT  2558..2879
FT  /*tag= f
FT  /label= 3'nos
FT  /note= "3' regulatory sequence contg. the
FT  polyadenylation signal of the nopaline
FT  synthase gene of Agrobacterium T-DNA"
FT  misc_RNA
FT  1..38
FT  /*tag= g
FT  /label= pUC19
FT  /note= "pUC19 derived sequence"
FT  2880..4986
FT  /*tag= h
FT  /label= pUC19
FT  /note= "pUC19 derived sequence"
XX  WO9534634-A2.
XX  21-DEC-1995.
XX
XX  06-JUN-1995; 95WO-EP02157.
XX
XX  06-JUN-1994; 94US-0254776.
XX
XX  (PLBZ ) PLANT GENETIC SYSTEMS NV.
XX
XX  Krebbers E, Leemans J, Williams M;
XX
XX  WPI; 1996-049664/05.
XX
XX  Transgenic plants contg. male sterility and colour-linked restorer
XX  genotypes - used for prodn. of male sterile seeds identifiable from
XX  their colour, also new truncated anthocyanin regulatory genes and
XX  aleurone specific promoters
XX
XX  Disclosure; Page 68-71; 104pp; English.
XX
XX  Shortened forms of anthocyanin (An) regulatory genes (ARG) are easier
CC

```

CC to manipulate than the complete gene and still provide an prodn.  
 CC The C1 gene (and the C1-S gene) can be considerably shortened  
 CC while still retaining, under appropriate conditions, its  
 CC capability of conditioning anthocyanin prodn. in the aleurone  
 CC of seeds of cereal plants such as corn. A pref. shortened C1 gene  
 CC is comprised in pCOL9 (see AAT08975). The full C1 gene sequence is  
 CC given in AAT08973.  
 CC A truncated B-peru gene (presumed sequence = AAT08674; actual  
 CC sequence = AAT08977) may also be used.  
 CC Plasmids pRS256 (AAT08976) and pRS200 (AAT17246) were used in the  
 CC construction of vectors comprising the C1 and B-peru genes  
 CC as well as male-sterility gene and a selectable marker gene.  
 CC  
 XX Sequence 4896 BP; 1252 A; 1146 C; 1186 G; 1312 T; 0 other;  
 SQ

Query Match 100.0%; Score 24; DB 17; Length 4896;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcagctccacc 24  
 |||||  
 Db 2316 tcagaagtatcagcagctccacc 2339

RESULT 13  
 AAX15631  
 ID AAX15631 standard; DNA; 4896 BP.  
 XX  
 AC AAX15631;  
 XX  
 DT 07-MAY-1999 (first entry)  
 XX  
 DE Nucleotide sequence of plasmid pRS256.  
 XX  
 KW C1 gene; maize; male-sterile corn line; anthocyanin production;  
 KW pRS256; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN US5860331-A.  
 XX  
 PD 09-MAR-1999.  
 XX  
 PF 07-JUN-1995; 95US-0485139.  
 XX  
 PR 07-JUN-1995; 95US-0485139.  
 PR 06-JUN-1994; 94US-0254776.  
 XX  
 PA (PLBZ ) PLANT GENETIC SYSTEMS NV.  
 XX  
 PI Krebbers E, Leemans J, Williams M;  
 XX  
 DR WPI; 1999-204053/17.  
 XX  
 PT Process for maintaining a male-sterile corn line - using male  
 PT sterile plants lacking functional regulatory gene for anthocyanin  
 PT production  
 XX  
 PS Example 2; Columns 31-36; 35pp; English.  
 XX  
 CC The present sequence represents the nucleotide sequence of plasmid  
 CC pRS256. It is used in the course of the invention. The specification  
 CC describes a process for maintaining a male-sterile corn line, using male  
 CC sterile parent plants lacking a functional gene for anthocyanin  
 CC production, and a maintainer corn line comprising male fertile parent  
 CC plants containing foreign DNA comprising a restorer gene and an active  
 CC regulatory protein gene. By using the anthocyanin gene, the colour of  
 CC the male-sterile plants will differ from that of the male fertile  
 CC plants. This will enable the seeds harvested from the plants to be  
 CC easily separated into those that will grow into male-fertile plants and  
 CC those that will grow into male-sterile plants.

SQ Sequence 4896 BP; 1251 A; 1147 C; 1186 G; 1312 T; 0 other;  
 Query Match 100.0%; Score 24; DB 20; Length 4896;  
 Best Local Similarity 100.0%; Pred. No. 0.069;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcagaagtatcagcagctccacc 24  
 |||||  
 Db 2317 tcagaagtatcagcagctccacc 2340

RESULT 14  
 AAT39339  
 ID AAT39339 standard; DNA; 5864 BP.  
 XX  
 AC AAT39339;  
 XX  
 DT 22-JAN-1997 (first entry)  
 XX  
 DE Plasmid pTCO113 T-DNA used to obtain male sterile oilseed rape.  
 DE Plasmid pTCO113; male sterile; barnase; ribonuclease; barstar;  
 KW transgenic plant; oilseed rape; canole; Brassica napus; ds.  
 KW  
 XX Synthetic.  
 OS  
 XX

Key Location/Qualifiers  
 FT misc\_feature complement (1..25)  
 FT /tag= a  
 FT /label= RB  
 FT /note= "right border of Agrobacterium T-DNA"  
 FT polyA\_signal complement (98..330)  
 FT /tag= b  
 FT /label= 3'g7  
 FT /note= "region containing polyA signal of gene 7  
 FT of Agrobacterium T-DNA"  
 FT CDS complement (331..882)  
 FT /tag= c  
 FT /label= bar  
 FT /note= "region coding for phosphinothricin  
 FT acetyltransferase"  
 FT promoter complement (883..2608)  
 FT /tag= d  
 FT /label= Psu  
 FT /note= "promoter of Arabidopsis Rubisco small  
 FT subunit gene"  
 FT polyA\_signal complement (2659..3031)  
 FT /tag= e  
 FT /label= 3'nos  
 FT /note= "region containing polyA signal of nopaline  
 FT synthase gene of Agrobacterium T-DNA"  
 FT CDS complement (3032..3367)  
 FT /tag= f  
 FT /label= Barnase  
 FT /note= "Bacillus amyloliquefaciens barnase coding  
 FT region"  
 FT promoter complement (3368..4877)  
 FT /tag= g  
 FT /label= pTA29  
 FT /note= "promoter of stamen-specific TA29 gene of  
 FT Nicotiana tabacum"  
 FT promoter 4924..5216  
 FT /tag= h  
 FT /label= Pnos  
 FT /note= "promoter of nopaline synthase gene of  
 FT Agrobacterium T-DNA"  
 FT CDS 5217..5489  
 FT /tag= i  
 FT /label= Barstar  
 FT /note= "region coding for barstar of Bacillus  
 FT amyloliquefaciens"  
 FT polyA\_signal 5490..5765

```
FT FT /*tag= j
FT FT /label= 3'g7
FT FT /note= "region containing polyA signal of gene 7
FT FT Of Agrobacterium T-DNA"
FT FT complement (5840..5864)
FT FT /*tag= k
FT FT /label= LB
FT FT /note= "left border of Agrobacterium T-DNA"
XX PN W09626283-A1.
XX PN
XX XX
XX XX
XX PD 29-AUG-1996.
XX XX
XX PF 21-FEB-1996; 96WO-EP00722.
XX XX
XX PR 21-FEB-1995; 95EP-0400364.
XX XX
XX PA (PLB2 ) PLANT GENETIC SYSTEMS NV.
XX XX
XX PI Botterman J, Cornelissen M, Michiels F;
XX XX
XX DR WPI; 1996-402373/40.
XX XX
XX PT Prodn. of male sterile plants by transforming with a chimaeric
FT construct - comprising a male sterility DNA e.g. barnase and a
FT co-regulating gene, e.g. barstar, into the nuclear genome, useful
FT for generating hybrid cultivars
XX XX
XX PS Example 3; Page 33-3743-47; 56pp; English.
XX XX
XX CC Plasmid pTCO113 (AMF39339) is a T-DNA vector containing a bar gene
CC under control of the P5S promoter, a barnase gene under control
CC of the stamen-specific pTA29 promoter, and a barstar (co-regulatory)
CC gene under control of the Pnos promoter. 87% Of oilseed rape
CC plants regenerated after Agrobacterium-mediated transformation
CC using pTCO113 were male sterile. Barnase expression disturbed the
CC function of stamen cells leading to male sterility. Constitutive
CC expression of barstar counteracted any low level expression of
CC barnase in non-stamen tissue.
XX XX
XX SQ Sequence 5864 BP; 1849 A; 1094 C; 1150 G; 1771 T; 0 other;

Query Match 100.0%; Score 24; DB 17; Length 5864;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
Db 5248 tcagaagtatcagcgacctccacc 5271

RESULT 15
AAD06990
ID AAD06990 standard; DNA; 5865 BP.
XX XX
XX AC AAD06990;
XX XX
XX DT 06-AUG-2001 (first entry)
XX XX
XX DE Chimeric T-DNA of plasmid pTCO113.
XX XX
XX KW T-DNA; plasmid pTCO113; transgenic Brassica plant; transformation event;
XX KW male-sterility gene; chimeric; tobacco; ds.
XX OS
XX OS Chimeric - Streptomyces hygroscopicus.
XX OS Chimeric - Arabidopsis thaliana.
XX OS Chimeric - Bacillus amyloliquefaciens.
XX OS Chimeric - Nicotiana tabacum.
XX OS Chimeric - Agrobacterium tumefaciens.
XX OS Chimeric - Unidentified.
XX XX
XX FH Key Location/Qualifiers
```

```
FT FT misc_feature
FT FT 1..25
FT FT /*tag= a
FT FT /note= "Right border repeat from the TL-DNA from
FT FT pTiB6S3"
FT FT 26..53
FT FT misc_feature
FT FT b
FT FT /*tag=
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 54..90
FT FT misc_feature
FT FT c
FT FT /*tag=
FT FT /note= "Residual sequence from the TL-DNA at the
FT FT right border repeat"
FT FT 98..309
FT FT 3'UTR
FT FT d
FT FT /*tag=
FT FT /note= "The 3' untranslated end from the TL-DNA
FT FT gene 7 (3'g7) of pTiB6S3"
FT FT 310..331
FT FT misc_feature
FT FT e
FT FT /*tag=
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 332..883
FT FT CDS
FT FT f
FT FT /*tag=
FT FT /product= "Protein encoded by bialaphos resistance
FT FT gene (bar) of Streptomyces hygroscopicus"
FT FT 884..2609
FT FT promoter
FT FT g
FT FT /*tag=
FT FT /note= "Promoter from the atSLA ribulose-1,5-biphosphate
FT FT carboxylase small subunit gene from Arabidopsis thaliana"
FT FT 2610..2659
FT FT misc_feature
FT FT h
FT FT /*tag=
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 2660..2920
FT FT misc_feature
FT FT i
FT FT /*tag=
FT FT /note= "TaqI fragment from the 3' untranslated end of the
FT FT nopaline synthase gene (3'nos) from the T-DNA of pTiT37
FT FT and containing plant polyadenylation signals"
FT FT 2921..2936
FT FT misc_feature
FT FT j
FT FT /*tag=
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 2937..3032
FT FT 3'UTR
FT FT k
FT FT /*tag=
FT FT /note= "The 3' untranslated region downstream from the
FT FT barnase coding sequence of Bacillus amyloliquefaciens"
FT FT 3033..3368
FT FT CDS
FT FT l
FT FT /*tag=
FT FT /product= "Protein encoded by barnase gene from
FT FT Bacillus amyloliquefaciens"
FT FT 3369..4878
FT FT promoter
FT FT m
FT FT /*tag=
FT FT /note= "Promoter region of the anther-specific gene
FT FT TA29 from Nicotiana tabacum"
FT FT 4879..4924
FT FT misc_feature
FT FT n
FT FT /*tag=
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 4925..5215
FT FT promoter
FT FT o
FT FT /*tag=
FT FT /note= "Promoter of the nopaline synthase gene from the
FT FT T-DNA of pTiT37 of Agrobacterium tumefaciens"
FT FT 5216..5217
FT FT misc_feature
FT FT p
FT FT /*tag=
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5218..5490
FT FT CDS
FT FT q
FT FT /*tag=
FT FT /product= "Protein encoded by barstar gene of
FT FT Bacillus amyloliquefaciens"
FT FT 5491..5530
FT FT misc_feature
FT FT r
FT FT /*tag=
FT FT /note= "Sequence from the 3' untranslated end of the
FT FT barstar gene from Bacillus amyloliquefaciens"
FT FT 5531..5554
FT FT misc_feature
FT FT s
FT FT /*tag=
FT FT /note= "Synthetic polylinker derived sequence"
FT FT 5555..5766
FT FT 3'UTR
FT FT t
FT FT /*tag=
FT FT /note= "The 3' untranslated end from the TL-DNA
```

```
FT      misc_feature      gene 7 (3'g7) of ptIB6S3"
FT      /tag= u
FT      /note= "Synthetic polylinker derived sequence"
FT      5767..5773
FT      /tag= u
FT      /note= "Synthetic polylinker derived sequence"
FT      5774..5810
FT      /tag= v
FT      /note= "Residual sequence from the TL-DNA at the
FT      right border repeat"
FT      5811..5840
FT      /tag= w
FT      /note= "Synthetic polylinker derived sequence"
FT      5841..5865
FT      /tag= x
FT      /note= "Left border repeat from the TL-DNA from
FT      ptIB6S3"
XX
PN      WO200131042-A2.
XX
XX      03-MAY-2001.
XX
XX      26-OCT-2000; 2000WO-EP10680.
XX
XX      29-OCT-1999; 99US-0430497.
XX
XX      (AVET ) AVENTIS CROPS SCIENCE NV.
XX
XX      Weston B, De Beuckeleer M;
XX      WPI; 2001-300517/31.
XX
XX      Transgenic Brassica plants, seeds, cells or tissues, characterized by
XX      harboring specific transformation events, particularly by presence of
XX      male-sterility gene, at specific location in its genome -
XX
XX      Claim 1; Page 47-49; 53pp; English.
XX
XX      The present invention relates to a transgenic Brassica plant or its
XX      seed, cells or tissues, characterised by harbouring a specific
XX      transformation event, particularly by the presence of a male-sterility
XX      gene, at a specific location in the Brassica genome. Transgenic
XX      Brassica plant is useful for producing a hybrid seed by crossing the
XX      transgenic plant with a male-fertile Brassica plant and harvesting the
XX      hybrid seed from the transgenic Brassica plant.
XX      The present sequence is chimeric T-DNA of plasmid pTCO113. This sequence
XX      comprises right border repeat, left border repeat and 3' untranslated
XX      region (UTR) from TL-DNA of ptIB6S3, synthetic polylinker sequences,
XX      coding regions of blalaphos resistance gene (bar) from
XX      Streptomyces hygroscopicus, barnase gene from Bacillus amyloliquefaciens
XX      and barstar gene from Bacillus amyloliquefaciens and promoters of atS1A
XX      ribulose-1,5-bisphosphate carboxylase small subunit gene from
XX      Arabidopsis thaliana, the anther-specific gene TA29 from
XX      Nicotiana tabacum and nopaline synthase gene from the T-DNA of pTi37
XX      of Agrobacterium tumefaciens.
XX
XX      Sequence 5865 BP; 1849 A; 1095 C; 1149 G; 1772 T; 0 other;

Query Match      100.0%; Score 24; DB 22; Length 5865;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      tcagaagtatcagcgacctccacc 24
          |||
Db      5249 tcagaagtatcagcgacctccacc 5272

Search completed: February 25, 2002, 18:17:36
Job time: 16694 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2002, 18:05:27 ; Search time 301.6 Seconds  
(without alignments)  
18.022 Million cell updates/sec

Title: US-09-698-903B-9

Perfect score: 24  
Sequence: 1 tcagaagtatcagcgacctccacc 24

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	1303	US-08-894-440-2	Sequence 2, Appli
C 2	24	100.0	3544	US-08-485-139-3	Sequence 3, Appli
C 3	24	100.0	3544	US-08-750-357-3	Sequence 3, Appli
C 4	24	100.0	4808	US-08-351-413-17	Sequence 17, Appli
C 5	24	100.0	4808	US-09-025-583-17	Sequence 17, Appli
6	24	100.0	4896	US-08-485-139-2	Sequence 2, Appli
7	24	100.0	4896	US-08-750-357-2	Sequence 2, Appli
8	24	100.0	5864	US-08-894-440-4	Sequence 4, Appli
C 9	24	100.0	6555	US-08-351-413-2	Sequence 2, Appli
C 10	24	100.0	6555	US-09-025-583-2	Sequence 2, Appli
11	16.6	69.2	1866	US-08-403-388-1	Sequence 1, Appli
12	16.6	69.2	1866	US-08-658-578-1	Sequence 1, Appli
13	16.6	69.2	1866	US-08-846-111D-1	Sequence 1, Appli
14	16.6	69.2	1866	US-09-036-105-19	Sequence 19, Appli
15	16.6	69.2	40352	US-08-846-111D-15	Sequence 15, Appli
C 16	16	66.7	1128	US-09-016-366A-20	Sequence 20, Appli
C 17	16	66.7	1128	US-08-978-404B-15	Sequence 15, Appli
C 18	16	66.7	1137	US-09-016-366A-18	Sequence 18, Appli
C 19	16	66.7	1137	US-08-978-404B-13	Sequence 13, Appli
C 20	16	66.7	1154	US-09-016-366A-16	Sequence 16, Appli
C 21	16	66.7	1154	US-08-978-404B-11	Sequence 11, Appli
C 22	15.6	65.0	4286	US-09-413-304-7	Sequence 7, Appli
C 23	15.6	65.0	4437	US-08-559-303B-72	Sequence 72, Appli
C 24	15.6	65.0	4437	US-09-175-828-72	Sequence 72, Appli
25	15.6	65.0	6519	US-08-588-985-1	Sequence 1, Appli
26	15.6	65.0	6519	US-08-971-988-1	Sequence 1, Appli
C 27	15.2	63.3	363	US-08-594-031-104	Sequence 104, App

c	28	15.2	63.3	1082	1	US-08-716-301-5	Sequence 5, Appli
	29	15.2	63.3	2326	2	US-08-231-193A-41	Sequence 41, Appli
	30	15.2	63.3	2326	2	US-08-486-273A-41	Sequence 41, Appli
	31	15.2	63.3	2326	3	US-08-480-474-41	Sequence 41, Appli
	32	15.2	63.3	2326	3	US-08-940-086A-41	Sequence 41, Appli
	33	15.2	63.3	3243	2	US-08-231-193A-44	Sequence 44, Appli
	34	15.2	63.3	3243	2	US-08-486-273A-44	Sequence 44, Appli
	35	15.2	63.3	3243	3	US-08-480-474-44	Sequence 44, Appli
	36	15.2	63.3	3243	3	US-08-940-086A-44	Sequence 44, Appli
	37	15.2	63.3	3698	2	US-08-231-193A-43	Sequence 43, Appli
	38	15.2	63.3	3698	2	US-08-486-273A-43	Sequence 43, Appli
	39	15.2	63.3	3698	3	US-08-480-474-43	Sequence 43, Appli
	40	15.2	63.3	3698	3	US-08-940-086A-43	Sequence 43, Appli
	41	15.2	63.3	3717	1	US-08-026-138E-7	Sequence 7, Appli
	42	15.2	63.3	4002	2	US-08-231-193A-53	Sequence 53, Appli
	43	15.2	63.3	4002	2	US-08-486-273A-53	Sequence 53, Appli
	44	15.2	63.3	4002	3	US-08-480-474-53	Sequence 53, Appli
	45	15.2	63.3	4002	3	US-08-940-086A-53	Sequence 53, Appli

ALIGNMENTS

RESULT 1  
US-08-894-440-2  
; Sequence 2, Application US/08894440  
; Patent NO. 6025546  
; GENERAL INFORMATION:  
; APPLICANT: PLANT GENETIC SYSTEMS N.V.  
; TITLE OF INVENTION: Method to obtain male sterile plants  
; FILE REFERENCE: NMSCOR  
; CURRENT APPLICATION NUMBER: US/08/894.440  
; NUMBER FILING DATE: 1997-11-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1303  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: HindIII-BcoRI  
; OTHER INFORMATION: fragment of pTS88  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(35)  
; OTHER INFORMATION: polylinker of pGEM2 (pGEM2)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (36)..(694)  
; OTHER INFORMATION: 35S promoter of Cauliflower Mosaic Virus strain  
; OTHER INFORMATION: CM1841 (p35S)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (695)..(967)  
; OTHER INFORMATION: region coding for barstar of Bacillus  
; OTHER INFORMATION: amyloliquefactions  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (968)..(1287)  
; OTHER INFORMATION: region containing polyadenylation signal of gene 7  
; OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1288)..(1303)  
; OTHER INFORMATION: polylinker of pGEM2  
; US-08-894-440-2

Query Match 100.0%; Score 24; DB 3; Length 1303;  
Best Local Similarity 100.0%; Pred. No. 0.0042;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 tcagaagtatcagcgacctccacc 24

```

Db 726 tcagaagtacagcgacctccacc 749
|||||
RESULT 2
US-08-485-139-3/C
; Sequence 3, Application US/08485139
; Patent No. 5880331
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,139
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3544 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII region of plasmid pTS200
; FEATURE:
; NAME/KEY: -
; LOCATION: 3227..3504
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2675..3226
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces
; OTHER INFORMATION: hygroscopicus"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1841..2674
; OTHER INFORMATION: /label= p35S
; OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (626..1803)
; OTHER INFORMATION: /label= PCA55
; OTHER INFORMATION: /note= "promoter of CA55 gene of Zea mays"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (353..625)

```

```

; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus
; OTHER INFORMATION: amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (30..352)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..6
; OTHER INFORMATION: /label= EcoRI
; FEATURE:
; NAME/KEY: -
; LOCATION: 3539..3544
; OTHER INFORMATION: /label= HindIII
; US-08-485-139-3
Query Match 100.0%; Score 24; DB 2; Length 3544;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcagaagtacagcgacctccacc 24
|||||
Db 594 TCAGAAGTATCAGCGACCTCCACC 571
|||||
RESULT 3
US-08-750-357-3/C
; Sequence 3, Application US/08750357
; Patent No. 6008437
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,357
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 018030-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3544 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: EcoRI-HindIII region of plasmid pTS200

```

```
FEATURE:
NAME/KEY: 3227..3504
LOCATION: 3227..3504
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: 2675..3226
LOCATION: 2675..3226
OTHER INFORMATION: /label= bar
OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces hygroscopicus"
FEATURE:
NAME/KEY: 1841..2674
LOCATION: 1841..2674
OTHER INFORMATION: /label= p35S
OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
FEATURE:
NAME/KEY: 626..1803
LOCATION: 626..1803
OTHER INFORMATION: /label= PCA55
OTHER INFORMATION: /note= "promoter of CA55 gene of Zea mays"
FEATURE:
NAME/KEY: 353..625
LOCATION: 353..625
OTHER INFORMATION: /label= barstar
OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus
OTHER INFORMATION: amyloliquefaciens"
FEATURE:
NAME/KEY: 30..352
LOCATION: 30..352
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: 1..6
LOCATION: 1..6
OTHER INFORMATION: /label= EcoRI
OTHER INFORMATION:
FEATURE:
NAME/KEY: 3539..3544
LOCATION: 3539..3544
OTHER INFORMATION: /label= HindIII
US-08-750-357-3

Query Match 100.0%; Score 24; DB 3; Length 3544;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24
Db 594 TCAGAAGTATCAGCGACCTCCACC 571

RESULT 4
US-08-351-413-17/c
; Sequence 17, Application US/08351413
; Patent No. 5750867
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/351,413
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/899,072
FILING DATE: 12-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,849
FILING DATE: 03-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 2121-102PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 4808 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: EcoRI-HindIII fragment of plasmid pYS218
FEATURE:
NAME/KEY: -
LOCATION: complement (18..401)
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the
OTHER INFORMATION: polyadenylation site derived from Agrobacterium
OTHER INFORMATION: T-DNA nopaline synthase gene"
FEATURE:
NAME/KEY: -
LOCATION: complement (402..737)
OTHER INFORMATION: /label= barnase
OTHER INFORMATION: /note= "coding region of the barnase gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: complement (738..1944)
OTHER INFORMATION: /label= pZM13
OTHER INFORMATION: /note= "promoter region of the Zm13 gene of Zea
OTHER INFORMATION: mays"
FEATURE:
NAME/KEY: -
LOCATION: complement (1945..2281)
OTHER INFORMATION: /label= 3'nos
FEATURE:
NAME/KEY: -
LOCATION: complement (2282..2554)
OTHER INFORMATION: /label= barstar
OTHER INFORMATION: /note= "coding region of the barstar gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: complement (2555..3099)
OTHER INFORMATION: /label= pTA29
OTHER INFORMATION: /note= "promoter region of the TA29 gene of
OTHER INFORMATION: Nicotiana tabacum"
FEATURE:
NAME/KEY: -
LOCATION: 3100..3932
OTHER INFORMATION: /label= 35S3
OTHER INFORMATION: /note= "35S3" promoter sequence derived from
OTHER INFORMATION: cauliflower mosaic virus isolate CabbB-J1"
FEATURE:
NAME/KEY: -
```

```

; LOCATION: 3933..4484
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of the phosphinothricin
; OTHER INFORMATION: acetyltransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4485..4763
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: 2333..2356
; OTHER INFORMATION: /label= BXOL2
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide
; OTHER INFORMATION: BXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2538..2586)
; OTHER INFORMATION: /label= PTA29SBOXL2
; OTHER INFORMATION: /note= "region complementary to oligonucleotide
; OTHER INFORMATION: PTA29SBOXL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2800..2823)
; OTHER INFORMATION: /label= PTA29OL5
; OTHER INFORMATION: /note= "region complementary to part of
; OTHER INFORMATION: oligonucleotide PTA29OL5"
; US-08-351-413-17
;
; Query Match 100.0%; Score 24; DB 1; Length 4808;
; Best Local Similarity 100.0%; Pred. No. 0.0054;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 tcgaagatcagcagcctccacc 24
; |||||
; Db 2529 TCAGAAGTATCAGCGCCTCCACC 2506
; |||||
;
; RESULT 5
; US-09-025-583-17/c
; Sequence 17, Application US/09025583
; Patent No. 5977433
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark
; APPLICANT: Leemans, Jan
; TITLE OF INVENTION: Maintenance of male-sterile plants
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,583
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.

```

```

; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: ECORI-HindIII fragment of plasmid pTS218
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (18..401)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (402..737)
; OTHER INFORMATION: /label= barnase
; OTHER INFORMATION: /note= "coding region of the barnase gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (738..1944)
; OTHER INFORMATION: /label= PZM13
; OTHER INFORMATION: /note= "promoter region of the 2m13 gene of Zea
; OTHER INFORMATION: mays"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (1945..2281)
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2282..2554)
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of the barstar gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2555..3099)
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "promoter region of the TA29 gene of
; OTHER INFORMATION: Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3100..3932
; OTHER INFORMATION: /label= 35S3
; OTHER INFORMATION: /note= "35S3" promoter sequence derived from
; OTHER INFORMATION: cauliflower mosaic virus isolate cabBB-J1"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3933..4484
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of the phosphinothricin
; OTHER INFORMATION: acetyltransferase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4485..4763
; OTHER INFORMATION: /label= 3'nos
; FEATURE:
; NAME/KEY: -
; LOCATION: 2333..2356
; OTHER INFORMATION: /label= BXOL2
; OTHER INFORMATION: /note= "region corresponding to oligonucleotide

```



```

; OTHER INFORMATION: BXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2538..2586)
; OTHER INFORMATION: /label= TA29SBXOL2
; OTHER INFORMATION: /note= "region complementary to oligonucleotide
; OTHER INFORMATION: TA29SBXOL2"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (2800..2823)
; OTHER INFORMATION: /label= PTA29OL5
; OTHER INFORMATION: /note= "region complementary to part of
; OTHER INFORMATION: oligonucleotide PTA29OL5"
; US-09-025-583-17
;
; Query Match 100.0%; Score 24; DB 2; Length 4808;
; Best Local Similarity 100.0%; Pred. No. 0.0054;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
Qy 1 tcagaagtatcagcgacctccacc 24
|||||
Db 2529 TCAGAAGTATCAGCGACCTCCACC 2506

RESULT 6
US-08-485-139-2
; Sequence 2, Application US/08485139
; Patent No. 5880331
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; NUMBER OF INVENTION: MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,139
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 010830-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: plasmid pTS256, linearized at HindIII
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (39..317)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
;
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (318..869)
; OTHER INFORMATION: /label= bar
; OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces
; OTHER INFORMATION: hygroscopicus"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (870..1702)
; OTHER INFORMATION: /label= P35S
; OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1740..2284
; OTHER INFORMATION: /label= PTA29
; OTHER INFORMATION: /note= "promoter of TA29 gene of Nicotiana
; OTHER INFORMATION: tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2285..2557
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of barstar gene of Bacillus
; OTHER INFORMATION: amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2558..2879
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation signal of the nopaline synthase
; OTHER INFORMATION: gene of Agrobacterium T-DNA"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..38
; OTHER INFORMATION: /label= pUC19
; OTHER INFORMATION: /note= "pUC19 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: 2880..4896
; OTHER INFORMATION: /label= pUC19
; OTHER INFORMATION: /note= "pUC19 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3004..3009
; OTHER INFORMATION: /label= EcoRI
; US-08-485-139-2
;
; Query Match 100.0%; Score 24; DB 2; Length 4896;
; Best Local Similarity 100.0%; Pred. No. 0.0054;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
Qy 1 tcagaagtatcagcgacctccacc 24
|||||
Db 2317 TCAGAAGTATCAGCGACCTCCACC 2340

RESULT 7
US-08-750-357-2
; Sequence 2, Application US/08750357
; Patent No. 6008437
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; NUMBER OF INVENTION: MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
```

```
STATE: Virginia
COUNTRY: United States
ZIP: 2213-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,357
FILING DATE: 21-MAR-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 018030-100
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4896 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: plasmid pTS256, linearized at HindIII
FEATURE:
NAME/KEY: -
LOCATION: 39..317
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: -
LOCATION: 318..869
OTHER INFORMATION: /label= bar
OTHER INFORMATION: /note= "coding region of bar gene of Streptomyces hygroscopicus"
FEATURE:
NAME/KEY: -
LOCATION: 870..1702
OTHER INFORMATION: /label= P35S
OTHER INFORMATION: /note= "35S promoter of Cauliflower Mosaic Virus"
FEATURE:
NAME/KEY: -
LOCATION: 1740..2284
OTHER INFORMATION: /label= PTA29
OTHER INFORMATION: /note= "promoter of TA29 gene of Nicotiana tabacum"
FEATURE:
NAME/KEY: -
LOCATION: 2285..2557
OTHER INFORMATION: /label= barstar
OTHER INFORMATION: /note= "coding region of barstar gene of
OTHER INFORMATION: Bacillus amyloliquefaciens"
FEATURE:
NAME/KEY: -
LOCATION: 2558..2879
OTHER INFORMATION: /label= 3'nos
OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation
OTHER INFORMATION: signal of the nopaline synthase gene of Agrobacterium T-DNA"
FEATURE:
NAME/KEY: -
LOCATION: 1..38
OTHER INFORMATION: /label= pUC19
OTHER INFORMATION: /note= "pUC19 derived sequence"
FEATURE:
NAME/KEY: -
LOCATION: 2880..4896
OTHER INFORMATION: /label= pUC19
OTHER INFORMATION: /note= "pUC19 derived sequence"
FEATURE:
NAME/KEY: -
```

```
LOCATION: 3004..3009
OTHER INFORMATION: /label= ECORI
US-08-750-357-2
```

```
Query Match 100.0%; Score 24; DB 3; Length 4896;
Best Local Similarity 100.0%; Pred. NO. 0.0054;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcagaagtatcagcgacctccacc 24
Db 2316 TCAGAGTATCAGCGACCTCCACC 2339
```

```
RESULT 8
US-08-894-440-4
Sequence 4, Application US/08894440
Patent No. 6025546
GENERAL INFORMATION:
APPLICANT: PLANT GENETIC SYSTEMS N.V.
TITLE OF INVENTION: Method to obtain male sterile plants
FILE REFERENCE: NMSCOR
CURRENT APPLICATION NUMBER: US/08/894,440
CURRENT FILING DATE: 1997-11-12
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 5864
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
OTHER INFORMATION: plasmid pTC0113
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement(1)..(25))
OTHER INFORMATION: right border of Agrobacterium T-DNA (RB)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((98)..(330))
OTHER INFORMATION: region containing polyadenylation signal of gene 7
OTHER INFORMATION: of Agrobacterium T-DNA (3'g7)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((331)..(882))
OTHER INFORMATION: region coding for phosphinothricin acetyl
OTHER INFORMATION: transferase (bar)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((883)..(2608))
OTHER INFORMATION: promoter of small subunit gene of Rubisco of
OTHER INFORMATION: Arabidopsis (Pssu)
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((2659)..(3031))
OTHER INFORMATION: region containing polyadenylation signal of
OTHER INFORMATION: nopaline synthase gene of Agrobacterium T-DNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((3032)..(3367))
OTHER INFORMATION: region coding for barnase of Bacillus
OTHER INFORMATION: amyloliquefaciens
FEATURE:
NAME/KEY: misc_feature
LOCATION: Complement((3368)..(4877))
OTHER INFORMATION: promoter of stamen-specific TA29 gene of Nicotiana
OTHER INFORMATION: tabacum (PTA29)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4924)..(5216)
OTHER INFORMATION: promoter of nopaline synthase gene of
OTHER INFORMATION: Agrobacterium T-DNA (Pnos)
```

FEATURE: NO  
NAME/KEY: misc\_feature  
LOCATION: (5217)..(5489)  
OTHER INFORMATION: region coding for barstar of Bacillus  
OTHER INFORMATION: amyloliquefaciens

FEATURE: NO  
NAME/KEY: misc\_feature  
LOCATION: (5490)..(5765)  
OTHER INFORMATION: region containing polyadenylation signal of gene 7  
OTHER INFORMATION: T-DNA (3'g7)  
FEATURE: NO  
NAME/KEY: misc\_feature  
LOCATION: Complement((5840)..(5864))  
OTHER INFORMATION: left border of Agrobacterium T-DNA  
US-08-894-440-4

Query Match 100.0%; Score 24; DB 3; Length 5864;  
Best Local Similarity 100.0%; Pred. No. 0.0056;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24  
|||||

Db 5248 tcagaagtatcagcgacctccacc 5271

## RESULT 9

US-08-351-413-2/c  
Sequence 2, Application US/08351413  
Patent No. 5750867

GENERAL INFORMATION:  
APPLICANT: Williams, Mark  
APPLICANT: Leemans, Jan  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA: US/08/351.413

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/899,072

FILING DATE: 12-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/370,849

FILING DATE: 03-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Svensson, Leonard R.

REGISTRATION NUMBER: 30,330

REFERENCE/DOCKET NUMBER: 2121-102PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 6555 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO  
ORIGINAL SOURCE: Plasmid pVE144 (replicable in E.coli)  
FEATURE: NO

NAME/KEY: NO  
LOCATION: 1..396  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"

NAME/KEY: NO  
LOCATION: complement (397..751)  
OTHER INFORMATION: /label= 3'nos  
OTHER INFORMATION: /note= "3' regulatory sequence containing the

OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA nopaline synthase gene"  
FEATURE: NO

NAME/KEY: NO  
LOCATION: complement (752..1024)  
OTHER INFORMATION: /label= barstar  
OTHER INFORMATION: /note= "coding region of the barstar gene of

OTHER INFORMATION: Bacillus amyloliquefaciens"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: complement (1025..1607)

OTHER INFORMATION: /label= TA29  
OTHER INFORMATION: /note= "promoter derived from the TA29 gene of  
OTHER INFORMATION: Nicotiana tabacum"

NAME/KEY: NO  
LOCATION: 1608..2440  
OTHER INFORMATION: /label= 35S3  
OTHER INFORMATION: /note= "35S3 promoter sequence derived from

OTHER INFORMATION: cauliflower mosaic virus isolate CabbB-JI"  
FEATURE: NO  
NAME/KEY: NO  
LOCATION: 2441..3256

OTHER INFORMATION: /label= neo  
OTHER INFORMATION: /note= "coding region of the neomycin  
OTHER INFORMATION: phosphotransferase gene of Tn5"

NAME/KEY: NO  
LOCATION: 3257..4315  
OTHER INFORMATION: /label= 3'ocs  
OTHER INFORMATION: /note= "3' regulatory sequence containing the

OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA octopine synthase gene"  
FEATURE: NO

NAME/KEY: NO  
LOCATION: 4316..6555  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"

US-08-351-413-2

Query Match

Best Local Similarity 100.0%; Score 24; DB 1; Length 6555;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcagaagtatcagcgacctccacc 24

|||||

Db 993 TCAGAAGTATCAGCGACCTCCACC 970

## RESULT 10

US-09-025-583-2/c

Sequence 2, Application US/09025583

Patent No. 5977433

GENERAL INFORMATION:

APPLICANT: Williams, Mark

APPLICANT: Leemans, Jan

TITLE OF INVENTION: Maintenance of male-sterile plants

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 2046
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; CURRENT APPLICATION DATA: US/09/025,583
; APPLICATION NUMBER: US/09/025,583
; FILING DATE:
; CLASSIFICATION: 800
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/351,413
; FILING DATE:
; APPLICATION NUMBER: US 07/899,072
; FILING DATE: 12-JUN-1992
;
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/970,849
; FILING DATE: 03-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 2121-102PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6555 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: plasmid pVcl44 (replicable in E.coli)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..396
; OTHER INFORMATION: /label= pUcl8
; OTHER INFORMATION: /note= "pUcl8 derived sequence"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (397..751)
; OTHER INFORMATION: /label= 3'nos
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA nopaline synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (752..1024)
; OTHER INFORMATION: /label= barstar
; OTHER INFORMATION: /note= "coding region of the barstar gene of
; OTHER INFORMATION: Bacillus amyloliquefaciens"
; FEATURE:
; NAME/KEY: -
; LOCATION: complement (1025..1607)
; OTHER INFORMATION: /label= TA29
; OTHER INFORMATION: /note= "promoter derived from the TA29 gene of
; OTHER INFORMATION: Nicotiana tabacum"
; FEATURE:
; NAME/KEY: -
; LOCATION: 1608..2440
; OTHER INFORMATION: /label= 35S3
; OTHER INFORMATION: /note= "35S3 promoter sequence derived from
; OTHER INFORMATION: cauliflower mosaic virus isolate Cabbb-JI"
; FEATURE:
;

```

```

; NAME/KEY: -
; LOCATION: 2441..3256
; OTHER INFORMATION: /label= neo
; OTHER INFORMATION: /note= "coding region of the neomycine
; OTHER INFORMATION: phosphotransferase gene of Tn5"
; FEATURE:
; NAME/KEY: -
; LOCATION: 3257..4315
; OTHER INFORMATION: /label= 3'ocs
; OTHER INFORMATION: /note= "3' regulatory sequence containing the
; OTHER INFORMATION: polyadenylation site derived from Agrobacterium
; OTHER INFORMATION: T-DNA octopine synthase gene"
; FEATURE:
; NAME/KEY: -
; LOCATION: 4316..6555
; OTHER INFORMATION: /label= pUcl8
; OTHER INFORMATION: /note= "pUcl8 derived sequence"
; US-09-025-583-2
;
; Query Match 100.0%; Score 24; DB 2; Length 6555;
; Best Local Similarity 100.0%; Pred. No. 0.0057;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; Qy 1 tcagaagtatcagcgacctccacc 24
; Db 993 TCAGAGTATCAGCGACCTCCACC 970
;
; RESULT 11
; US-08-403-388-1
; Sequence 1, Application US/08403388
; Patent No. 5587289
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-Xp Family and Uses Thereof
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,388
; FILING DATE: 14-MARCH-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587289man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1866 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-403-388-1
;
; Query Match 69.2%; Score 16.6; DB 1; Length 1866;
; Best Local Similarity 82.6%; Pred. No. 21;
; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| | |||||
DB 498 CAGAACCTCAGGAGCTCCACC 520

RESULT 12
US-08-658-578-1
; Sequence 1, Application US/08658578
; Patent No. 5759783
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-Xp Family and Uses Thereof
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,578
; FILING DATE: 5-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,388
; FILING DATE: 14-MARCH-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5759783man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5444
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1866 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-658-578-1

Query Match 69.2%; Score 16.6; DB 1; Length 1866;
Best Local Similarity 82.6%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| | |||||
DB 498 CAGAACCTCAGGAGCTCCACC 520

RESULT 13
US-08-846-111D-1
; Sequence 1, Application US/08846111D
; Patent No. 6017705
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-B Family and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York

```

```

; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect For DOS 6.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,111D
; FILING DATE: 25-APRIL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/658,578
; FILING DATE: 5-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,388
; FILING DATE: 14-MARCH-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6017705man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5444.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1866 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-846-111D-1

Query Match 69.2%; Score 16.6; DB 3; Length 1866;
Best Local Similarity 82.6%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| | |||||
DB 498 CAGAACCTCAGGAGCTCCACC 520

RESULT 14
US-09-056-105-19
; Sequence 19, Application US/09056105
; Patent No. 6287569
; GENERAL INFORMATION:
; APPLICANT: KIPPS, THOMAS J.
; APPLICANT: WU, YUNQI
; TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR
; TITLE OF INVENTION: PROCESSING
; FILE REFERENCE: 233/221
; CURRENT APPLICATION NUMBER: US/09/056,105
; CURRENT FILING DATE: 1998-04-06
; EARLIER APPLICATION NUMBER: 60/043,467
; EARLIER FILING DATE: 1997-04-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 1866
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-056-105-19

Query Match 69.2%; Score 16.6; DB 4; Length 1866;
Best Local Similarity 82.6%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cagaagtatcagcgacctccacc 24
   ||||| |||| | |||||
DB 498 cagaagctcaggagctccacc 520

```

```

RESULT 15
US-08-846-111D-15
; Sequence 15, Application US/08846111D
; Patent No. 6017705
; GENERAL INFORMATION:
; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
; TITLE OF INVENTION: Members Of The Mage-B Family and Uses Thereof
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect For DOS 6.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,111D
; FILING DATE: 25-APRIL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/658,578
; FILING DATE: 5-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,388
; FILING DATE: 14-MARCH-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6017705man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LOD 5444.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40352 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-846-111D-15

```

```

Query Match          69.2%; Score 16.6; DB 3; Length 40352;
Best Local Similarity 82.6%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cagaagtatcagcgacctccacc 24
||||| ||||| ||||| |||||
Db 38322 CAGAGGCTCAGGAGCTCCACC 38344

```

Search completed: February 25, 2002, 18:05:29  
Job time: 18577 sec